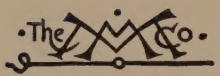
STUDIES IN SERVICE AND SELF-CONTROL

HARTSHORNE AND MAY

R. L. V. Leonhardt
July 1931

STUDIES IN SERVICE AND SELF-CONTROL



THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO

MACMILLAN & CO., LIMITED
LONDON · BOMBAY · CALCUTTA
MELBOURNE

THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

STUDIES IN THE NATURE OF CHARACTER

BY THE

CHARACTER EDUCATION INQUIRY

TEACHERS COLLEGE, COLUMBIA UNIVERSITY

IN COÖPERATION WITH

THE INSTITUTE OF SOCIAL AND RELIGIOUS RESEARCH

II

STUDIES IN SERVICE AND SELF-CONTROL

HUGH HARTSHORNE, MARK A. MAY, AND JULIUS B. MALLER

BOOK ONE

STUDIES IN SERVICE

BOOK TWO

STUDIES IN SELF-CONTROL

United Theological Seminary
LC
268
.C75
1928a
V. 2
41,830

New York

THE MACMILLAN COMPANY

1929

14977
The Evangelical
School of Theology

COPYRIGHT, 1929,
BY THE MACMILLAN COMPANY.

All rights reserved, including the
right of reproduction in whole
or in part in any form.

Set up and electrotyped. Published October, 1929.

• PRINTED IN THE UNITED STATES OF AMERICA •

FOREWORD

THE steps leading to the inauguration of the Character Education Inquiry were detailed in the Foreword to *Studies in Deceit*, the first volume issuing from the Inquiry. It will therefore suffice in this second volume to recount only the salient facts.

The Religious Education Association and other national bodies had urged on the Institute of Social and Religious Research the desirability of a searching study into the outcomes of customary religious and ethical instruction. After prolonged consultation with specialists, the Institute in 1924 arranged to have such a study adopted by Teachers College, Columbia University, the funds being supplied by the Institute; and Dr. Hugh Hartshorne and Dr. Mark A. May were engaged as codirectors. A group of advisers was brought together to counsel with the investigators during the progress of the Inquiry; and it was on its recommendation that the period of study was extended from the initial three years to five years, ending August 31, 1929.

To lay minds this volume, at first glance, may seem overloaded with matter that has little to do with moral and religious education — a medley of tests and statistics and a paucity of clear directions as to building character. Such readers might profitably reflect that these preliminary processes are inevitable if character education is ever to emerge from guesswork into a science. Medical and surgical science had to follow a similar road to advance from magic and quackery. Instead of continuing to treat the body as a mysterious unit, the biologists subjected each organ and constituent to minute observation and controlled experimentation. In like manner, it was at length recognized that we could gain little exact knowledge about character and how to change it unless its component features were accurately studied and experimented with.

One question before the investigators in the Inquiry was to select as units for study such modes of conduct as were not only

UNITED THEOLOGICAL SEMINARY

significant in living but also susceptible of precise measurement and experimentation. They answered this question by deciding to focus their attention on deceitful and honest behavior, as reported in *Studies in Deceit*, and on coöperative and charitable behavior and on self-control, as reported in this volume.

It must be left to time and the experts to pass judgment on the daring work done by Professors May and Hartshorne. But it may not be out of place to observe that they have not only supplied scientific evidence for doubting the efficacy of certain conventional efforts at character development in both secular and religious circles, but have also contributed to the more difficult task of building anew the whole structure of moral and religious education.

GALEN M. FISHER, Executive Secretary
Institute of Social and Religious Research

ACKNOWLEDGMENTS

Dr. E. L. Thorndike, under whose general supervision the Inquiry has been conducted, has not only read the manuscript of the report and made valuable suggestions regarding its details, but has also afforded constant encouragement and wise counsel. Dr. Ella Woodyard has read both manuscript and proof, for which we are very grateful.

We are particularly indebted to the Board of Education, Superintendent of Schools, the teachers, and the parents of Walden, New York, for their splendid coöperation during the year 1927–28, when our extensive test program was being administered. The four schools in New Haven which carried part or all of our test program showed equal forbearance and consideration. We asked a great deal of the teachers and principals, and they spent many out-of-school hours supplying us with needed information about their pupils.

Finally, we are under obligation to the Department of Education at Yale University for additional space in which to carry on certain phases of the work and to several members of its staff for their courtesy and assistance.

During the preparation of this volume of our report, our own staff has included as appointees of Teachers College the following persons: Helen Barthel, Chester A. Kerr, Julius B. Maller, Randolph B. Smith.

In addition, Miss Ruth E. Welty worked out the individual tests of self-control, and Captain C. E. Christenberry was primarily responsible for the self-control "parties."

H. H.
M. A. M.

New York,
March 1, 1929.

CONTENTS

CHAPTER PAGE

INTRODUCTION 1

The General Method of Inquiry. Subject Matter of Volume II. Populations Utilized. General Conclusions

BOOK ONE

STUDIES IN SERVICE

PART I

THE MEASUREMENT OF COÖPERATIVE AND CHARITABLE BEHAVIOR

I. PRELIMINARY EFFORTS TO STUDY ALTRUISTIC TENDENCIES . 13

The Human Background of Unselfishness. Analysis of the First Attempts. Production Techniques. Self-Sacrifice Techniques. Validation of the Self-Sacrifice Techniques. Scaling Self-Sacrifice. The Relation of Preliminary Measures of Helpful Behavior to Miscellaneous Facts

II. ANALYSIS OF SERVICE AS AN OBJECT OF STUDY . . . 42

Behavior Scales Originating in Conflict. The Significance of Motive, Conflict, Competition, and Coöperation. Previous Efforts to Measure Charity and Coöperation

III. THE CEI SERVICE TESTS 53

The Self-or-Class Test. The Money Voting Test. The Learning Exercises. The School Kit Test. The Envelopes Test. The Efficiency Coöperation Test. The Free Choice Test. The Total Service Scores. Summary

IV. REPUTATION FOR SERVICE 74

The Record of Social Service. The Portrait Matching Device. The "Guess Who" Test. The Check List. Total Reputation for Service

V. THE VALUE OF THE SERVICE TESTS 98

Descriptive Résumé. The Reliability of the Service Tests. The Validity of the Service Tests. Conduct and Reputation. Theoretical Validity. True Validity. Conclusions

PART II

FACTORS ASSOCIATED WITH SERVICE

CHAPTER PAGE

VI. Age, School Status, and Intelligence 119
Age and Service. Grade and Service. Retardation and Acceleration. Intelligence and Service. Scholastic Marks and Service. Conclusions concerning Age, Grade, and Intelligence

VII. Sex Differences in Coöperative Tendencies 146
Sex Differences in Test Scores. Sex Differences in Reputation Scores. Grade and Sex. Conclusions

VIII. Physical and Emotional Condition in Relation to the Tendency to Be of Service 158
Physical Condition. Emotional Condition

IX. General Environment as a Factor in Effective Service . 169
Community Differences. Population Differences in Service. Occupational Differences. Economic and Cultural Differences. National Differences. Religious Differences. Conclusions

X. The Influence of the Home on the Tendency to Be of Service 200
Sibling Resemblance in Service. Parental Coöperation and the Service Scores of Children. Parental Intelligence and the Service Scores of Children. Conclusion

XI. The Specific Character of School Groups 208
School Morale. Classroom Morale. Conclusions concerning School and Classroom Morale

XII. Out-of-School Experience as a Factor in Service . . 218
Sunday-School Attendance. Systems X and Y and Service. Summer Camps and Service. Motion Picture Attendance and Service

XIII. Miscellaneous Factors Associated with Service Tendencies 233
Suggestibility and Service. Cleanliness. Sociability. Friendship. The Effect of Majorities on Service. The Effect of Teams on Coöperative Activity. The Effect of the Repetition of Service Tests. The Instability of the Motive of Coöperation. Biological Types and Coöperative Behavior. The Specificity of Service Tendencies. The Effect of Teaching on Coöperation

XIV. Characteristics of Contrasting Groups 252
Comparative Significance of Factors Associated with Service. The Prevalence and Strength of Unselfishness.

PART III

CONCLUSIONS AND IMPLICATIONS

CHAPTER PAGE

XV. Conclusions and Implications of Studies in Service . . 263
Results of Primary Studies. Results of Secondary Studies. Implications

BOOK TWO

STUDIES IN SELF-CONTROL

PART I

THE MEASUREMENT OF SELF-CONTROL

XVI. Analysis of Self-Control 279
Intermingling of Persistence and Inhibition. The Function of Interest. Varieties of Interest Used in the Tests. Self-Control and Character.

XVII. Previous Attempts to Measure Inhibition and Persistence 285
Classification of Tests

XVIII. Classroom Tests for Measuring Persistence and Inhibition 292
Persistence. Inhibition. The Total Self-Control Scores. Administration of Self-Control Tests

XIX. Reputation for Self-Control 321
Reputation for Persistence. Reputation for Inhibition

XX. Value of the Classroom Tests of Self-Control . . 328
The Persistence Tests: Reliability; Correspondence between Conduct and Reputation; Theoretical Validity; True Validity. The Inhibition Tests: Reliability; Correspondence between Conduct and Reputation; Theoretical Validity; True Validity. Conclusions concerning the Value of the CEI Measures of Self-Control

XXI. Individual Tests of Self-Control 342
Persistence. Inhibition. Scoring the Individual Tests. Value of the Individual Tests

XXII. Tests of Inhibition in Party Games 353
The General Plan. Description of Tests. Reliability of the Tests. Validity of the Tests. Results of the Party Tests

PART II

FACTORS ASSOCIATED WITH SELF-CONTROL

CHAPTER PAGE

XXIII. Age, School Status, and Intelligence . . . 361

Age and Self-Control. Grade Status and Self-Control. Intelligence and Self-Control. Teachers' Marks and Pupils' Self-Control. Supplementary Discussion of Interest and Self-Control. Conclusions concerning Age, Grade, and Intelligence

XXIV. Sex and Physical and Emotional Condition in Relation to Self-Control 380

Sex. Possible Explanation of Sex Differences in Self-Control. Physical Condition. Emotional Condition.

XXV. General Environment as a Factor in Self-Control . 390

Community Differences. Summary of Community Differences in Self-Control. Occupational Differences. Economic and Cultural Differences. Summary of Occupational and Cultural Differences in Self-Control. National Differences. Religious Differences. Summary of National and Religious Differences in Self-Control. Conclusions concerning General Environment as a Factor in Self-Control.

XXVI. Home and School in Relation to Self-Control . . 407

The Home. The School. Conclusions concerning Home and School in Relation to Self-Control

XXVII. Out-of-School Experience as a Factor in Self-Control 418

Sunday Schools. Conclusions concerning the Relation between Sunday-School Attendance and Self-Control. System Y. Conclusions concerning Clubs and Self-Control. Summer Camps. Motion Pictures. Conclusions concerning Motion Picture Attendance and Self-Control

XXVIII. Miscellaneous Factors Associated with Self-Control 428

Suggestibility and Self-Control. Cleanliness. Sociability and Self-Control. Friendship and Self-Control. Speed Ability and Self-Control. The Effect of Majorities on Self-Control. Characteristics of Contrasting Groups. The Specificity of Self-Control

PART III

CONCLUSIONS AND IMPLICATIONS

CHAPTER PAGE

XXIX. Conclusions and Implications of Studies in Self-Control 443

Results of Primary Studies. Results of Secondary Studies. Implications

Appendices 455

Index 553

TABLES

BOOK ONE

(The tables of Appendix C are listed on pages 466–470)

TABLE		PAGE
I.	Comparison of Helpers and Non-Helpers in One Class	22
II.	Comparative Helpfulness of X and Non-X Children	23
III.	Intercorrelation of Helpful Acts in a Suburban Community (A)	28
IV.	Intercorrelation of Helpful Acts in a Private School (P)	28
V.	Correlations between Actual Gifts and Amounts Voted to Self or Charity and to Bank	33
VI.	Intercorrelations of Self-Sacrifice Tests	37
VII.	Correlations between Self-Sacrifice Scores and Scaled Scores	38
VIII.	Correlations between Self-Sacrifice Scores and Various Facts	40
IX.	Scoring Table for Money Vote	59
X.	Score Values of Articles in Kit Test	63
XI.	Score Values of Envelopes Test for Populations X, Y, and Z	67
XII.	Evaluation of Original Ten Portraits	81
XIII.	Distribution of Scores on Portraits — Population P	86
XIV.	Reliability of Portrait Matching — Population P	87
XV.	Effect of Signing One's Name on a Ballot	89
XVI.	Distribution of "Guess Who" Scores — Population P	90
XVII.	Intercorrelations of Measures of Reputation for Service	96
XVIII.	Correlation of Each of the Service Reputation Scores with the Sum of the Other Two — Population P	97
XIX.	Correlations between Service Tests and the Total Reputation for Service — Population P	103
XX.	Correlation of Service Tests with the Service Reputation — Population XYZ	104
XXI.	Sex Differences in r's between Service and Reputation	104
XXII.	Intercorrelations of Preliminary Service Tests — Population P	111
XXIII.	Intercorrelations of Final Service Tests — Population XYZ	111
XXIV.	Correlations between Age and the Separate Service Tests	123
XXV.	Percentage Distribution of Grade Status by Years for Each Population and Grade	132

TABLE		PAGE
XXVI.	Percentage Distribution of All Cases by Age and Grade	133
XXVII.	Percentage Distributions of Normal, Accelerated, and Retarded Groups by Grade and Population	134
XXVIII.	Basic Physical Ability Scores	159
XXIX.	Comparison of Median Scores on the Woodworth-Mathews Test with the Mathews Tentative Norms	164
XXX.	Summary of Population Differences	176
XXXI.	Correlations between 22 Classroom Averages on Burdick and Service Tests	187
XXXII.	Correlation between Siblings in the Service Tests	202
XXXIII.	Length of Attendance of Each Age-Group of Each School	209
XXXIV.	Certain Group Differences in Total Service Scores	213
XXXV.	Rank Difference Correlations of the Class Service Scores of Boys with Those of Girls — Populations X, Y, and Z	215
XXXVI.	Resemblance of Friends and Classmates in Total Service Score and IER School Xi	237
XXXVII.	Correlation of Mutual Friends of Same Class in Service Scores	238
XXXVIII.	Individual Variability of Work for Self and Work for Group	242
XXXIX.	Probability of Occurrence of Differences between Selfish and Unselfish Groups	258

BOOK TWO

XL.	Typical Illustrations of Average Number of Examples Done for Self and for Class in the First Two and Last Two Periods of the Efficiency Coöperation Test	301
XLI.	Relative Attractiveness of Objects in the Puzzle Manipulation Test	307
XLII.	Scoring Key for the Puzzle Manipulation Test	309
XLIII — A.	Standardization Data for the Ruggles Test	312
XLIII — B.	Standardization Data for the Ruggles Test (*Continued*)	313
XLIV.	Reliabilities of Persistence Tests	330
XLV.	Correlations between Tests and Reputations for Persistence	331
XLVI.	Intercorrelation of Persistence Tests — Population XYZ	332
XLVII.	Correlations between Inhibition Tests and Reputation	336
XLVIII.	Intercorrelation of Inhibition Tests — Population XYZ	339
XLIX.	Intercorrelation of Inhibition Tests — Population C	340
L.	Distributions of Time Scores of Individual Persistence Tests	349

TABLE		PAGE
LI.	Intercorrelations of Individual Inhibition Tests	351
LII.	Validity of Inhibition Tests as Measured by the Sum of Six	351
LIII.	Validity of Inhibition Tests as Measured by the Self-Control Ratings	352
LIV.	Reliabilities of the Party Inhibition Tests	357
LV.	Inter-*r*'s of Party Inhibition Tests	358
LVI.	Correlations between Persistence Tests and Age — Populations X, Y, and Z	363
LVII.	Correlations between Intelligence (CAVI Sigma) Scores and Persistence Scores — Populations X, Y, and Z	373
LVIII.	Correlations between Intelligence (CAVI Sigma) Scores and Inhibition Scores — Populations X, Y, and Z	374
LIX.	Correlations between Siblings in the Persistence Tests	408
LX.	Correlations between Siblings in the Inhibition Tests	409
LXI.	Correlations between Self-Control Medians of Boys and Girls — All Classes in Populations X, Y, and Z	416
LXII.	Resemblance of Friends of Same Classroom in Self-Control	436
LXIII.	Resemblance of Friends of Different Classes in Self-Control	437

FIGURES

BOOK ONE

FIGURE		PAGE
1.	Results of Toy-Making Experiment	26
2.	Disposition of Money Gift by Vote and Cash	30
3.	Percentages Taken from Amounts Voted to Self, Bank, and Charity When Gift Dropped from 25¢ to 19¢	31
4.	Proportion Giving All to Self, All to Charity, and Part to Both Self and Charity	31
5.	Changeability of Givers and Non-Givers	32
6.	Disposition of Money Gift on First and Second Votes	32
7.	The Scaling of Helpful Behavior	39
8.	Imaginary Behavior Scales with Zero Point at a State of Social Conflict	44
9.	Illustration of Learning Exercises	58
10.	Facsimile of One Side of Sheet Used for the Coöperation Tests	69
11.	Percentage Distribution of the Service Reputation Scores — All Populations Combined	105
12.	Correlation Table of Service Scores and Reputation Scores — All Populations Combined	108
13.	Percentage Distribution of Total Service Scores — All Populations Combined	109
14.	Median Total Service Scores by Age Compared with the Median for All — Populations X, Y, and Z	122
15.	Per Cent at Each Age Coöperating on Three Service Tests — Population Z	124
16.	Median Total Service Scores by Grade Compared with the Median for All — Populations X, Y, and Z	126
17.	Mean Scores by Grade on Efficiency Coöperation and Money Vote Tests — Population Z	128
18.	Per Cent of Each Grade Coöperating on Three Service Tests — Population Z	129
19.	Median Service Scores of Normal, Accelerated, and Retarded Pupils of Population Y	135
20.	Mean Service Reputation Scores of Accelerated, Normal, and Retarded Boys and Girls — Population Y	137
21.	Comparison of Bright, Average, and Dull Pupils with Regard to Median Service Scores — Populations X, Y, and Z	140

FIGURE		PAGE
22.	Comparison of Very Helpful, Average, and Selfish Pupils with Regard to Median Intelligence Scores — Populations X, Y, and Z	141
23.	Per Cent of Each Sex Coöperating on the Free Choice Test — Populations X, Y, and Z	147
24.	Sex Differences on the Free Choice Test	148
25.	Distribution of Free Choice Scores by Sex	149
26.	Sex Differences during the Efficiency Coöperation Test	150
27.	Sex Differences in the Average Number of Examples per Minute Done for Self and for Class in Each Two-Minute Period of the Efficiency Coöperation Test	151
28.	Per Cent of Each Sex Coöperating on the Envelopes Test	153
29.	Sex Differences in Reputation Scores — Populations X, Y, and Z Combined	155
30.	Condition of Nutrition and of Teeth in Relation to Median Service Score	162
31.	Percentage Distribution of Neurotic Index Scores Based on Pupil Data Sheets	166
32.	Comparison of Service Scores of Nail Biters and Non-Nail Biters	168
33.	Percentage Distribution of Burdick Culture Scores — Populations X, Y, and Z	173
34.	Percentage Distribution of Sims Score for Populations X and Y Only	174
35.	Percentage Distribution of Occupational Levels — Populations X, Y, and Z	175
36.	Percentage Distribution of Total Service Scores by Population	178
37.	Population Differences in Median Scores of the Free Choice and Efficiency Coöperation Tests, and in Median Scores on the Money Vote Test	179
38.	Population Differences in Per Cent Coöperating on the Kits and Envelopes Tests — Populations P, X, Y, and Z	180
39.	Occupational Differences in Median Service Score — Populations X, Y, and Z	182
40.	Occupational Differences in Two Service Tests — Populations X and Z	183
41.	Contrast between Boys and Girls in Occupational Differences on the Money Vote and Kits Tests — Population Z	184
42.	Cultural Differences in Median Scores — Populations X, Y, and Z	186
43.	Scattergram of Classroom Means of Sims Score Card and Total Service Scores for Populations X and Y	188
44.	Comparative Service Scores of Children of Native and Foreign-born Parents	190

FIGURE PAGE
45. Median Service Scores of Children One or Both of Whose Parents Were Born in Ireland or Italy 191
46. National Differences in Median Culture and Socio-economic Scores 193
47. Comparison of Median Service Scores of Children Whose Parents Are Affiliated with the Three Religious Groups — Populations X, Y, and Z 195
48. Comparison of Mean Culture Scores of Children Whose Parents Are Affiliated with the Three Religious Groups — Populations X, Y, and Z 196
49. Comparison of Mean Sims Scores of Children Whose Parents Are Affiliated with the Three Religious Groups — Populations X and Y 197
50. Comparison of Scores on Separate Service Tests of Children Whose Parents are Affiliated with the Three Religious Groups — Population X 198
51. Median Service Scores According to Length of Attendance at Each School — Populations X, Y, and Z 211
52. Sunday-School Attendance in Relation to Median Service Scores for Selected Groups in Population Y 220
53. Sunday-School Enrollment and Median Service Scores — Populations X and Z 221
54. Sunday-School Attendance in Relation to Median Service Scores — Populations X and Z 222
55. Median Service Scores of Members and Non-Members of Organizations — Populations X and Y 226
56. Comparison of Members and Non-Members of Organizations in Reputation for Service — Populations X and Y 227
57. Movie Attendance in Relation to Service — Population X . . 229
58. Movie Attendance in Relation to Service — Population Z . . 230
59. Movie Attendance in Relation to Total Service Medians — Population Y 231
60. Median Service Scores for Different Degrees of Cleanliness — Population Y 235
61. Effect of Fatigue on Work Done for Self and for Class — Boys 244
62. Percentage Distribution of Service Reputation Scores of Selfish and Unselfish Pupils — Population XYZ 253
63. Percentage Distribution of Ages of Selfish and Unselfish Pupils — Population XYZ 254
64. Proportionate Representation of National Groups among the Selfish and Unselfish Pupils — Populations X, Y, and Z . 255
65. Proportionate Representation of Religious Groups among the Selfish and Unselfish Pupils — Populations X, Y, and Z . . 256

FIGURE PAGE

66. Proportionate Representation of Organizations (System Y) among the Selfish and Unselfish Pupils — Populations X and Y 257

BOOK TWO

67. Magic Number Square or P-M 1 Test 298
68. Magic Word Square or P-M 2 Test 299
69. Facsimile of the Ruggles Distraction Test 308
70. Illustration of the Picture Inhibition Test 315
71. Percentage Distribution of the Inhibition Reputation Scores — All Populations Combined 337
72. Percentage Distribution of the Total Inhibition Scores — All Populations Combined 338
73. Per Cent Failing to Inhibit on Each Trial of Each Party Test . 359
74. Percentage Distributions of Hop Race Scores and Total Party Inhibition Scores 360
75. Median Total Persistence Scores by Age Compared with the Median for All — Populations X, Y, and Z 362
76. Comparative Frequency of Inhibition among Children Ten to Eleven and Children Twelve to Thirteen — Population XYZ 366
77. Median Total Persistence Scores by Grade Compared with the Median for All — Populations X, Y, and Z 368
78. Median Total Inhibition Scores by Grade Compared with the Median for All — Populations X, Y, and Z 369
79. Mean Persistence Scores of Accelerated, Normal, and Retarded Pupils of Population Y 370
80. Mean Inhibition Scores on Three Tests of Accelerated, Normal, and Retarded Boys and Girls of Population Y 372
81. Median Persistence Scores of Bright, Average, and Dull Pupils of Population Y 373
82. Sex Differences in Mean Persistence Score and Mean Reputation Score — Populations X, Y, and Z 381
83. Sex Differences in Mean Total Inhibition Score (upper) and Mean Reputation Score (lower) — Populations X, Y, and Z . . 383
84. Vision, Nutrition, and Condition of Teeth in Relation to Persistence 386
85. Median Neurotic Scores for Various Levels of Persistence — Population XYZ 387
86. Median Neurotic Scores for Various Levels of Scores on Three Inhibition Tests — Population YZ 388
87. Percentage Distribution of Total Persistence Scores by Population 391

FIGURE PAGE
88. Population Differences in Median Scores on the Cross and Ring, Magic Square, and Story Resistance Tests 392
89. Population Differences in Median Scores on the Persistence for Self and for Class Tests 392
90. Population Differences in Conduct on the Four Inhibition Tests 394
91. Comparative Persistence Scores of Children of Native and Foreign-born Parents — Populations X, Y, and Z 398
92. Comparative Scores in Reputation for Self-Control of Children of Native and Foreign-born Parents — Populations X, Y, and Z 400
93. Comparative Inhibition Scores of Children of Native and Foreign-born Parents — Populations X, Y, and Z 401
94. Comparison of Median Persistence Scores of Children Whose Parents Are Affiliated with the Three Religious Groups — Populations X, Y, and Z 403
95. Comparison of Median Inhibition Scores of Children Whose Parents Are Affiliated with the Three Religious Groups — Populations X, Y, and Z 404
96. Median Persistence Scores According to Length of Attendance at Each School — Populations X, Y, and Z 413
97. Sunday-School Attendance in Relation to Median Persistence Scores for Selected Groups in Population Y 419
98. Mean Persistence Scores of Members and Non-Members of Organizations — Populations X and Y 422
99. Mean Inhibition Scores of Members and Non-Members of Organizations — Populations X, Y, and Z 424
100. Movie Attendance in Relation to Total Persistence Medians — Populations X, Y, and Z 426
101. Movie Attendance in Relation to Total Inhibition Medians — Populations X, Y, and Z 427
102. Effect of Social Suggestion on Persistence of Pupils . . . 431
103. Effect of Social Suggestion on Persistence of Pupils (*Continued*) 432
104. Median Self-Control Scores for Different Degrees of Cleanliness — Population Y 434

INTRODUCTION

STUDIES IN SERVICE AND SELF-CONTROL

INTRODUCTION

THE GENERAL METHOD OF INQUIRY

In reporting the findings of an extensive and complex research the temptation is strong to try to say everything all at once. If it were possible to present in one picture the entire set of facts which have been brought to light in the Character Education Inquiry, the result would doubtless conform more nearly to the realities from which our data are drawn. Such complex facts, however, constitute the common experience of humanity and have been thus confronted in their totality for centuries. New insights into their significance are to be gained by the method not of literary realism but of analytic science. It is the details of the picture, which we have been unable to see because of our preoccupation with life in its concrete wholeness, that we need now to examine in order that we may view human nature in all its dimensions, as it were, rather than merely in its surfaces. The human body is the same as it was a thousand years ago. But what a totally different object it seems to us from what it did to our ancestors! Knowledge of such facts as the circulation of the blood, the processes of digestion, the nature of infection, and the physiology of the nervous system makes an extraordinary difference in our viewpoint. Science, by filling in the details of the picture, is providing us not only with a new perspective but also with new avenues of control. So, with those even more subtle aspects of human nature we call mind and character, we move toward more perfect understanding and control by the processes of analysis and experiment. In mastering himself, man has found it of great advantage to know how he is made and how his various parts work together.

In psychology there are many fruitful modes of analysis, and human nature is so complex that every possible way of studying it should be welcomed. Our own approach, which was described in the introduction to the first volume of our report, *Studies in Deceit*, requires the isolation of certain types of behavior, so far as these are exhibited in intelligent response to certain types of situations, for the purpose of discovering functional tendencies and their interrelations. In schematic form, the objects of our study described in Volume I are as follows:

1. Mental contents and skills — the so-called intellectual factors
2. Desires, opinions, attitudes, motives — the so-called dynamic factors
3. Social behavior — the performance factors
4. Self-control — the relation of these factors to one another and to social-self-integration

The fourth item represents the concrete reality which we hope to understand better from our study of the first three groups of facts.

As was indicated in Volume I, such an investigation as the Character Education Inquiry can hope to study only a few samples of these various elements. Beginning with the third item, we selected a typical social failure — the tendency to deceive — as one unit for detailed analysis and attempted to deal with it objectively and in quantitative units. Having measured the amount of deceptiveness characterizing individuals in a variety of situations, we were able to determine the relation between this tendency and other measurable or countable facts, such as age, intelligence, schooling, suggestibility, sex, and emotional stability.

SUBJECT MATTER OF VOLUME II

The present volume sets forth the methods and results of our investigation of two other types of behavior, each of which, as in the case of deceit, ranges theoretically from a condition of unsatisfactory adjustment to one of satisfactory adjustment. The first couplet, which we have for convenience called "service," contrasts work for self and work for others. The second couplet, for which

we now use the term "self-control" in a special sense, contrasts the tendency to continue an approved act with resistance to the tendency to engage in an interesting but disapproved act. As we proceed various names will be attached to these behaviors, and they will be described in detail. We may refer to them tentatively as coöperation, charity, persistence, and inhibition.

It is quite obvious that these tendencies are closely related and interwoven. Almost any act involves all of them. Nevertheless, they are distinguishable both in motive and overt response. Paradoxically, a child may be helpful for selfish motives. He may perform acts which are no less of service to others in being undertaken to gain his own ends. On the other hand, with all the good will in the world a bungler may make a nuisance of himself in trying to help someone else. Or the essence of helpfulness may at times be the willingness to keep everlastingly at it, as when one attempts to resuscitate a person who has been long submerged. To persist in this act of helpfulness would require a rigid inhibition of tendencies to rest, to eat, to engage in other activities. From one point of view, therefore, we may be said to be studying motives, and from another point of view, behaviors.

As in our studies of deception, we shall attempt to deal primarily with the behaviors; but it must be recognized that in making this abstraction for purposes of investigation we are not belittling the significance of motives, but are merely insisting on the necessity of knowing the tendency, under any motive, for the overt responses to be made in certain situations. The problem of motivation is not irrelevant, but it is another problem.

This piecemeal method of studying character, in which we take up one type of behavior at a time, requires that we present the material of the present volume in several distinct sections. We shall deal first with what we have called service and shall discuss persistence only in so far as the techniques for measuring the tendencies to work for self and for others also measure persistence in these activities. The second section, Book Two, will deal with those aspects of self-control which we have called persistence and inhibition. In each book, Part I will describe and appraise the

techniques we have developed, Part II will report the associations found between the ability to score on our tests and such concomitant factors as could be measured or counted, and Part III will summarize these results and suggest certain implications for moral education.

As no technical devices will be employed which have not been fully described in *Studies in Deceit*, we shall omit all but the briefest explanations of statistical terms and methods and shall place all technical material in footnotes, small type, and appendices. It is suggested that those to whom the language of statistical inquiry is somewhat unfamiliar would find it helpful to read again Chapters V and VI of Book One, *Studies in Deceit*, or any elementary text in statistics.

The general reader is cautioned against regarding the techniques to be described as tests of character. We do not consider them to be anything more than tests of specific forms of behavior. Their value as elements in a test of character as a whole will be reported in Volume III.

POPULATIONS UTILIZED

In working out the techniques reported in the present volume, we used four of the populations in which we had already done considerable work. We shall designate these by the letters applied to them in *Studies in Deceit*.

A. A suburban community of 10,000 population with about 1000 children in the public school grades 5 to 8, with many national and religious groups and a wide social range. There are five elementary schools, grades 1 to 6, one intermediate school, grades 7 to 8, and one high school.

B. A public school serving a metropolitan population which on the whole is above the average in social and economic conditions. The average intelligence quotient is considerably above normal.

D. Another group of public school children, who because of broken homes are situated in an institution for such chil-

dren located in a suburban community. This is a rather unique group in that it represents the low end of the scale in the way of home background and economic level. The average intelligence of this group is less than normal.

P. A private school, grades 1 to 6, having mixed sexes, and grades 7 and 8, only girls, located in a large city and drawing from the upper social levels. The average intelligence quotient is well above 100.

In addition we had available a population of nearly 900 children which we were studying intensively with our entire battery of tests. In this volume we shall refer only to those aspects of this program which have direct bearing on service and self-control (in the narrow sense). These groups we shall call X, Y, and Z.

X. Grades 5 and 6 of a public school in a residential section of New Haven, Connecticut. The community is not wealthy, but might be called upper middle class. X also contains grades 7 and 8 of another school, likewise representing a community of relatively high social level.

Y. This group was selected because of its representative character. The entire population in grades 5 to 8 is included. The town, known as Walden, contains 5000 people and is situated about ten miles due west of Newburgh, New York The inhabitants are largely of old English stock.

Z. Grades 5 to 8 from a third school in New Haven, Connecticut. The community is largely foreign, but with no colored or oriental pupils. Three-fourths of the fathers are unskilled laborers.

We thus have a mid-group, Y, a high group, X, and a low group, Z.

GENERAL CONCLUSIONS

Our studies in service and self-control have resulted in a large number of techniques. We have made use of the results of fourteen of these, involving some twenty-six tests in all. This is

a reasonably satisfactory sample, since it contains about one-sixth of the total number of techniques that a complete study of these types of conduct would require.

Our chief purpose in the inquiry has been to explore the field of character measurement and to demonstrate the feasibility and value of quantitative procedures in this type of research. The feasibility of such procedures depends on their cost in time and money and their effect on the pupils. We have no reason to suppose that the tests used have done the pupils any harm, and it is our opinion that the activities and choices involved contributed to their well-being.

The cost in time and money, including the administration and scoring of tests, will seem great or small in proportion to the results. But it can be stated quite definitely that the measurement of trends of conduct takes more of the pupils' time than is now spent on the measurement of mental abilities and achievements. Samples of conduct, however, could be obtained in short testing periods, which would give a reliable measure of the particular type of conduct involved in the particular situation presented; and by this means individuals and populations could be compared, and the results of experiments could be measured. But a broad view of individual or group conduct would require a large number of such samples.

The value of these procedures depends on what may be learned by their use As already suggested, one may compare individuals and populations and one may learn the results of methods which are used to change conduct. We have undertaken a variety of sample studies of this nature, and something of what we have learned from them will now be briefly summarized to illustrate the possibilities of this type of research.

Studies in Service. We find that with children in grades 5 to 8 service tendencies are specific rather than general and are made more in response to the external demands of the situation than to inner views of the principles or ideals involved. Generosity is about as prevalent as greed, and group loyalty is usually as strong a motive as the desire for reward or recognition. The friends a child

goes with, the classroom code, satisfactory school adjustment, the example of parents, and general cultural level are all more or less tied up with tendencies to be of service; but age, intelligence, sex, emotional condition, the power to resist suggestion, and sociability are only loosely connected with coöperation and charity. Communities, national groups, and religious groups are factors of importance in determining the helpful behavior of children. Sunday-school and club members are slightly more coöperative than non-members, as is the case also with those who go to motion picture shows less frequently than their schoolmates.

The tendency of pupils to help one another can be changed, at least temporarily, by suitable teaching.

Studies in Self-Control. We find also that self-control, like the other types of conduct so far studied, is specific and is functionally related to the situations to which the responses are made. Children cannot be divided into two groups, the well controlled and the badly controlled: most of them are moderately controlled.

Age, intelligence, nervous stability, and power to resist suggestion are all more or less associated with persistence; but culture and economic status make only a slight difference. Friends and siblings tend to resemble one another, and girls are a little more persistent than boys. The scores of the pupils depend a great deal on the community in which they live and on their national background; but these differences are specific, tending in one direction on some tests and in the opposite direction on others. Length of Sunday-school attendance is slightly correlated with greater persistence, but on the whole those enrolled in Sunday schools are less persistent on some tests than those not enrolled, presumably because the latter belong to those population groups which do well on these tests. The scores of club members are not distinguishable from those of non-members; but the day-school classroom is an influential factor, largely, no doubt, because the tests are given in classrooms.

The facts for inhibition are somewhat different. Here sex is a highly significant factor, the girls proving far better inhibited than the boys. This may be due to the greater interest of girls in school

grades and their fainter interest in the manipulation of mechanical puzzles and the like. Age, intelligence, emotional condition, suggestibility, and home background make very little difference in inhibition scores; but friends and siblings are somewhat more alike than would be two children chosen at random from the whole school. Communities and national groups differ considerably and in opposite directions on the various tests, with no community or group consistently superior on all. Sunday-school or club membership makes no difference; but the classroom, probably because the tests are classroom tests, proves to be a significant factor in the situation.

Implications. The tendencies to be of service to others, to exercise self-restraint, and to overcome obstacles or fatigue are learned just like any other skill. Children do not develop *general* tendencies in these directions, but specific tendencies according to the experiences they have had. These types of conduct are largely matters of morale. They have to be learned piecemeal, but they are supported by the group code. When morale breaks down, the associated habits of self-control tend to go also.

There is little evidence that the children tested have been subjected to an organized program of moral education, if we accept consistently socialized conduct as a natural outcome of such education. While it is true and encouraging that the children dealt with have been neither ungenerous nor uncontrolled, it is also true that their conduct trends seem to be the accidental products of fragmentary experience rather than the planned results of well-organized experience.

The teaching of coöperation, charity, and self-control requires (1) careful planning of situations to which these activities are the natural and successful response, (2) provision for building a group morale which supports the desirable mode of conduct, and (3) increasing complexity and difficulty of situations in order that general principles may emerge and be brought into play for the guidance of conduct and the integration of behavior.

The bearing of these problems of specificity and integration on character as a whole will be taken up in Volume III.

BOOK ONE

Studies in Service

PART I

The Measurement of Coöperative and Charitable Behavior

CHAPTER I

PRELIMINARY EFFORTS TO STUDY ALTRUISTIC TENDENCIES

THE HUMAN BACKGROUND OF UNSELFISHNESS

There seems to be a feeling abroad that children are far more selfish than adults. There is also a strong impression that one need not expect much in the way of group loyalty and coöperation from the young members of the human family until they are adolescent. It is not clear just how the transition from interest in self to interest in others takes place or just what causes the individualist of seven to turn into the loyal teamworker of sixteen. There are occasions in the experience of most families when the apparent ingratitude and selfishness of children is appalling, but there are few parents who ever raise the question as to the origin of this behavior. It is taken for granted that the youngsters were born that way and must have the old Adam either gently or forcibly removed.

But there are also occasions in the experience of some families — probably in most — when the children rise to heights of self-sacrifice rarely equaled by adults. It is hard for parents to see things through the eyes of children, to understand their longings and affections. We excuse ourselves for ignoring their disappointments by saying that laughter easily follows tears. But the

poignancy of frustration is theirs perhaps in sharper measure for all its ready transition to other moods, and when a child loses or gives away a cherished object, he loses or gives away an intimate part of himself which can never be wholly replaced.

Observant parents, also, may find even in the five-year-old such staunch loyalty to his family as it might well behoove his father or mother to exhibit more frequently (though probably in wiser ways). A couple of youngsters of this age were calling the passers-by all sorts of names, such as "you old garbage can," when the older of the two started to apply these epithets to the younger boy's mother, whereupon the latter launched a hot attack, both verbal and physical, upon his bigger playmate. And as for unselfishness, there are many parents who could duplicate in their experience the sacrifice of the little daughter who had been saving all year to buy a pair of skates, but who gave all her money for the relief of Russian children when the need came to her attention. What would be an equivalent gift for an adult?

All this is the human background of our studies in service. If we forsake the picture of real children in a real world to deal with experiments and statistical analyses, in the language of uncolored scientific statement, it is in order to attempt to understand more clearly the nature and causes of those powers of coöperation and self-sacrifice which are the very foundation of an enduring and happy society. Unfortunately, the ages of the subjects with which we are working do not allow us to confirm or reject the popular opinions expressed in the first paragraph, but at least we are able to challenge their accuracy. As a possible substitute, we offer the opinion that children vary enormously among themselves in their tendencies to be of service. The aggregate of habits peculiar to the child, which we label selfish or generous as the case may be, is learned in the course of his experience with others.

ANALYSIS OF THE FIRST ATTEMPTS

In 1925 to 1926 we started our first experiments in the measurement of helpful tendencies. As the analysis of the problem which

we made at that time was used in guiding our early attempts at devising techniques, we reproduce it here:

1. Variables involved
   (*a*) Stimulus to helpful behavior
   (*b*) Resistance to helpful behavior
2. Types of stimuli
   (*a*) Hospital children of various social groups
   Orphans
   Poorer children
   Children relatively well off
   Children of different races and religions
   (*b*) Younger children of various social groups as under (*a*)
   (*c*) Children of the same age — various groups
   (*d*) Older children — various groups
   (*e*) Adults — various groups
   (*f*) Old people — various groups
   (*g*) Animals — various groups
   (*h*) Institutions and causes — graded as to nearness to subject's experience
   (*i*) The community
3. Types of resistance
   (*a*) Things to be sacrificed
   Money — for candy, movies, etc.
   Food — particularly sweets
   Toys
   Articles made for oneself
   Time — for movies, reading, games, sleep, etc.
   (*b*) Things to be done
   Making or doing things (standardized articles or acts, as toys, mending books)
   Writing letters
   Making visits
   Collecting things
   Earning money by doing things (standardized activities, such as correcting other children's tests)

As both the stimulus and the resistance are operating to determine the behavior, a series of tests should keep one variable constant while changing the other. That is, it would be necessary to

select one stimulus, such as hospital children of one or more social groups or younger children in the same school, and keep this stimulus the same while varying the resistances to be overcome in making a helpful response, using as many resistances as possible for each situation. For example, opportunity might be offered to give either money or fruit to hospital children, or time saved from games might be devoted to making toys or picture books for them or in writing them letters. Theoretically, the amount of the tendency to help — no matter what the motive — is measured by the amount of resistance overcome in helping.

Limiting Conditions. In attempting to work out an experiment along the lines just described we were confronted with certain public school conditions which definitely limit what an outside agency can accomplish. These limitations are as follows:

1. If all the subjects to be measured are to have the same opportunity to be of help, whatever is to be done has to be done at school. But the free hours of the school session are few in number and different for different pupils. Since such hours are usually study periods, the motive of escape from study, varying greatly among the pupils, would confuse the issue.

It had been suggested that some school be asked to change its routine temporarily so as to provide certain free hours for all, with the understanding that the pupils who wished to work on the project to be proposed would make up their work at home at the sacrifice of play hours. A school system might well do this, but we did not feel like imposing to this extent on the good nature of the principals coöperating with us, nor did we wish to take the responsibility of restricting the outdoor play hours of the more helpfully inclined children.

2. The use of schoolrooms out of school hours, requiring no free-period adjustments, means difficulties from the standpoint of the parents as well as loss of out-of-door play. After-school hours would require additional janitor service and the responsibility of teachers and principals for buildings and property. The half hour before school opens, however, avoids most of these problems, and its use was attempted, as will presently be described.

3. Earning money during school hours even on free time was felt to be objectionable, and the after-school hours were unavailable for the reasons just now given.

4. The teachers cannot easily be eliminated from the situation in spite of all precautions and explanations. In the case of the experiment using the half hour before school opened, some teachers had other plans for the use of their time. Others were so much interested that they could not keep quiet about the project and unconsciously influenced the children to come early. Some opposed the plan, claiming it made pupils tired, and so discouraged attendance.

Standards of Procedure. Under the circumstances, it was difficult to hold ourselves rigidly to scientific standards of measurement. The criteria of a good test for measuring helpful tendencies are comparable to those we outlined for test building in our *Studies in Deceit.* The following paragraphs repeat these standards in a form adapted to the situations with which we are now concerned.

1. The test situation should be as far as possible a natural situation. It should also be a controlled situation. The response should as far as possible be natural even when directed.

2. The test situation and the response should be of such a nature as to allow all subjects equal opportunity to exhibit the behavior that is being tested. That is, there should be nothing about the test situation itself which would prevent anyone who desired to coöperate from doing so; on the other hand, there should be nothing about it to persuade a person who did not wish to help to do so for selfish reasons.

3. No test should make use of appeals or present situations calling for help that are unusual or extreme.

4. The tests should make use of *genuine* situations. The subject should not be deceived into responding to an imaginary situation presented as real.

5. The tests should have "low visibility"; that is, they should be of such a nature as not to arouse the suspicions of the subject.

6. If the tests are to be used in statistical studies, they should be group tests. They should be easy to administer and should be mechanically scored. They should be short enough to be given in single school periods.

7. The test results should be clear and unambiguous. It should be

obvious from the results whether the subject did or did not exhibit the behavior in question.

8. The score should be quantitative, showing the amount as well as the fact of service. Each test should therefore be flexible enough to include within its scope wide ranges of coöperative tendency.

As in the case of our deception techniques, we were unable to meet all these requirements in any one test, but they served as a valuable set of standards in the selection and refinement of the techniques for measuring service to others. Reference to our criteria, for example, enabled us to weed out many apparently good suggestions without actual trial. One of the most attractive proposals made was to have the subjects score a number of short-answer tests that had been administered to younger children, either to help the teacher with the work or in order to earn money with which to participate in a stated social service project. The amount and quality of work done would be a measure of the tendency to coöperate in the project. But no natural setting for such a large undertaking could be found which would offer papers enough for scoring, particularly since the helpful aspect of the act would be greatly reduced by the necessity of rescoring. We were unwilling to create an artificial situation, as indicated in requirement number four.

We were also obliged to abandon an interesting project which called for the making of large quantities of cardboard objects, such as protractors or rulers, for the use of younger children. Again our reason was that we found the demand for such objects was not adequate for the supply that would result from any large-scale experiment. An adaptation of this idea, however, was adopted, which will presently be described.

The Plans Tried Out. The plans actually tried out may be outlined as follows:

1. Production techniques
   (*a*) Picture pasting
   (*b*) Toy making
   (*c*) Participation in social service activities

2. Self-sacrifice techniques
   (*a*) Giving up money
   (*b*) Giving up ice cream

In the remainder of this chapter we shall describe these devices and present such results as seem to be significant.

PRODUCTION TECHNIQUES

Picture Pasting

We secured the coöperation of the principal of school C and of a local hospital for this experiment. From the hospital a letter was addressed to the pupils as follows:

To the Children of P. S. ——:

We have many little patients in our children's wards who find the long hours very tiresome. We want to make these hours pass quickly and happily for them, for they miss their homes and their own toys and playthings.

Now children love pictures, and even sick children can look at pictures if they are mounted on light cards. But after a sick child once uses a picture card it has to be burned. That means that a great many cards are always needed.

So we thought that you might like to help us make a lot of picture cards. We have arranged to provide the material if you will do the work.

You may be sure that whatever you care to do will make many little sick children very happy.

(Signed)

The day before the experiment began with any class, the examiner went to the class in the afternoon and spoke as follows:

We have a letter from the hospital which I want to read to you at this time. (Reads letter.) All of you who would like to make picture cards for these children may come to-morrow morning at ten minutes past eight.

As we shall have only a half hour to work, we want to make as many cards as we can and so we hope none of you who plan to come will be late.

Only those who are on time or nearly so will be allowed to work on the cards. I have the slips for you to have your parents sign to give you permission to come. Remember, we start at 8 : 10.

The examiner then gave out school excuse slips, asking the parents' written consent to the pupils' coming to school early. Any children who wished could take the slips. The rest returned them.

The next morning at 8 : 10 the examiner appeared in the classroom, with cardboards for mounting the pictures, a copy of the *Saturday Evening Post* of a certain date, from which pictures could be cut, and scissors and paste. The directions for the classroom procedure were as follows :

Directions for Making Cards

Examiner should have a seating plan of the children of the class. Keep the time when each child arrives up to fifteen minutes, after which time let no child arriving work.

Examiner should be in schoolroom in time to have scissors, spatula, two cardboards, and magazine on each child's desk in readiness so that no time will be lost. As each child arrives, supply with paste.

"We are going to make these cards for the children in the hospitals and we want to make them as attractive as we can."

Examiner takes care to avoid attracting any child to herself. The work is not to be made interesting for any extraneous reason.

The name and grade must be written on each card completed and handed to examiner, who puts it out of sight to avoid children's learning results. When half hour is up, the names should be written on unfinished work and all material gathered up and put away for the next day. Children must not work on cards during the day.

(When the rest of the children arrive at beginning of school), "Perhaps some of you were here yesterday. We are making picture cards for the little children in the hospital. If any more of you would like to come, I shall be here at ten minutes past eight."

Take all completed cards to home office. Do not mention to the children how many have responded to work.

After the first day, each child was allowed to bring pictures from home, and this was the general practice. In spite of all efforts to

keep the teacher out of the situation, we found that by encouragement or the reverse the attendance was materially affected by what the teachers said to the children.

Results of the Picture-Pasting Technique. The picture-pasting experiment was tried out in four classes, $5A^2$, $5B^2$, $6A^2$, $6A^3$. The last group was a bright class, to which we had already given a large number of tests of various kinds. As already stated, the hospital letter was read to each class on the day before they were to come to school early; and on the following morning an examiner was on hand at eight o'clock in each room to receive the children and set them to work. Of the $5B^2$ group only one child, a boy, came the first morning, and after three mornings he dropped out. The $6A^3$ class was started on a Wednesday and the other two on the next Friday. The work ceased on the following Friday. In the case of the $6A^3$ group, a helpfulness score was worked out on the basis of frequency and promptness of attendance and amount of work accomplished. Credit was also given for extra pictures brought from home.

The resistance was too high to make this a practical test. Only about 39% of $6A^3$ came at least once, 32% of $6A^2$, and 46% of $5A^2$. Those who did come were well measured, however, as only seven out of fifty came *every* morning, and of these only one was always on time. The scores of $6A^3$ were distributed in percentages as follows:

Scores	Per Cent of Class
0–9	68
10–19	8
20–29	12
30–39	4
40–49	0
50–59	2
60–69	2
70–79	0
Over	4

This table reads: 68% of the class got scores between 0 and 9; 8% got scores between 10 and 19; etc.

The number attending various lengths of time in the case of the two groups $6A^2$ and $5A^2$ may be of some interest:

Number of Days at Work	Number of Children
0	50
1	9
2	5
3	5
4	4
5	5
6	4

Those in $6A^3$ had eight days in all, and those in the other two classes had six days. The project was then called off on account of the fact that the teachers' influence was increasingly making itself felt, sometimes for and sometimes against the early attendance of the children. The pupils who came were often eager to keep on and to bring extra material from home. A few children did considerable additional work outside and brought it in.

It is of interest to inquire whether there were any striking differences between the children who came early to school at least once in order to make the picture cards and those who did not come at all. Such facts as are available for the $6A^3$ group are gathered in Table I.

TABLE I

Comparison of Helpers and Non-Helpers in One Class

	% of 28 Boys	% of 21 Girls	CA	IQ	Sims Score*
Helping . . .	32	43	123.6	144.8	14.1
Not helping . .	68	57	127.4	143.1	12.5

* For a description of the Sims Score Card see *Studies in Deceit*, Book One, Chapter VIII. The Card is published by the Public School Publishing Company. It gives a numerical rating representing general socio-economic level.

From Table I it can be seen that a somewhat smaller proportion of boys attended than of girls. Those who came were also a little younger and of higher IQ. These two differences are not significant, but the helpers have a decidedly higher home background score as measured by the Sims Score Sheet.

In view of the fact that in *Studies in Deceit* we had occasion to contrast the members of a certain organization called system X with the non-members in the matter of deception, it may be of interest to report the facts regarding helpfulness. It will be recalled that we made no claim to have measured the general usefulness of this system of character training and pointed out the necessity for a more comprehensive measure of its product. It happened that the X records of 41 of the 49 children in the 6A[3] group were available. Of these 41 pupils, 25 were X's and 16 were non-X's. The numbers are too few for generalization, but comparisons are printed in Table II for what they may be worth.

TABLE II

COMPARATIVE HELPFULNESS OF X AND NON-X CHILDREN

	% OF X'S	% OF NON-X'S	MEAN H SCORE OF X'S	MEAN H SCORE OF NON-X'S	MEAN RANK IN X
Helping . . .	52	25	34.8	32.8	2.8
Not helping . .	48	75			2.4

While about as many X's came as stayed away, the proportion of all the X's who came was considerably greater than the proportion of the non-X's who came — 52% as against 25%. The average helpfulness score of the X group was also slightly higher than that of the non-X group. That is, they came a little more regularly and accomplished somewhat more.

There were four possible ranks in the X organization for these children, 1, 2, 3, and 4. The X's who helped had evidently been slightly more active than the non-helpers in filling out the report cards on which advancement depended, as their average rank is .4 of a point higher than that of those who did nothing to help.

Whether these same differences would appear in other classrooms is, of course, wholly unknown.

All the cards furnished by the pupils as well as the wooden ducks which the boys made in the toy-making experiment were sent to the hospital, and the note of thanks which was sent in reply was read to all who had taken part.

Toy Making

A letter similar to the one used for the pictures was used as the occasion for helpful behavior. This time the letter asked for toys instead of pictures. By previous arrangement with the shop director of school C, one-half of a 7B class of boys was chosen for this experiment. In their regular shop work, these boys had carried along a project of their own (making a large wooden automobile) to the point where all were deeply interested in it. This constituted the uniform resistance against which the situation of sick hospital children needing toys was presented. This appeal was as nearly alike for a group of children as anything we could imagine. While the task had interest, it was rather meager as compared with the making of the automobile. Consequently any variations in practice might justly be attributed to differences in the attitude of helpfulness.

Only one group was employed for this technique, as we found the boys were very slow in making the toys, taking as much as three hours to finish one toy. Each boy of course made his own. The director kept at work on the automobile with the boys who preferred to do that. The examiner was always present to be of assistance and to keep track of results. His directions were as follows:

Directions for Toy Making

Grade 7B[1]

Cause — Hospital children

On Thursday read to this grade a letter from the hospital requesting wooden toys to be made for the sick children.

Activity — Making wooden toys

Time — Friday mornings only — 9 : 20 to 10 : 20 (examiner to arrive about 9 : 10 to have everything in readiness)

Equipment — Furnished by Inquiry: Saws, pattern (duck)
Furnished by school: Wood, carbon, styli, sandpaper, and hardware needed

Keep a record of all work done by each pupil and of the time each boy works. Bring all completed toys to Room 300, Teachers College. Procedure to continue as long as children work. Let no boys work outside of this hour.

The choice of some object which could be more quickly made would improve this test materially and make it quite usable.

Results of the Toy-Making Technique. The few cases make a graphic presentation of results rather necessary. Figure 1 shows the behavior of each pupil. The blanks are absences. The shaded sections begin where the subjects left the toy-making project and returned to the automobile. The double hatching shows days at work, and the black spaces indicate the completion of a duck. The amount of double hatching is supposed to show the relative potency of the unselfish interest.

The first six toys were handed in unpainted. The second set of toys were painted on the boys' own initiative.

As the shop period came only once a week, the time consumed in this experiment was too extensive to make it practicable as a test.

Participation in Social Service Activities

In a suburban community where we had been doing a good deal of testing (population A) and also in a local private school (population P) we found that the teachers knew with reasonable certainty the relative amount of help given by the children in several service enterprises carried on during the year. As this information was quantitative in character, we took steps to have it recorded. Each teacher was provided with a list of her pupils and the objects to which the pupils had been giving their assistance. The directions which follow were used in making the records:

Directions for Recording Acts of Service

We are making a study of why some children are more helpful in certain situations than other children. Probably no test of helpfulness that could

be given in a brief time would be as valuable as the actual experiences of helpfulness through which your school has been passing. We have put at the top of the page the names of certain enterprises common to your

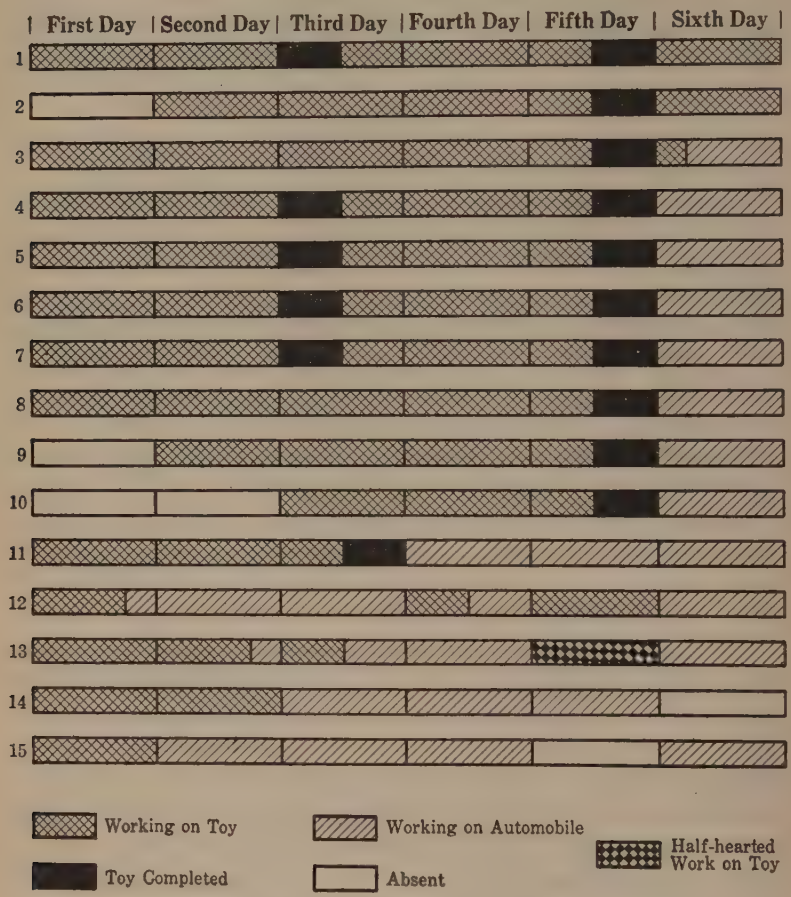


FIGURE 1

RESULTS OF TOY-MAKING EXPERIMENT

Each row shows the work of one boy. Thus boy 1 finished a toy by the third day and then started another, which he finished on the fifth day, when he started a third, which he worked on through the sixth day.

community. Probably there are others your own class has engaged in. *Please add* these other enterprises in the vacant columns *if these pupils* engaged in them.

Please make the record as follows: Take one column at a time and work down through the class. Obviously the children did different amounts of work. These amounts are represented by the scale 0, 1, 2, 3, 4.

0 means he did nothing at all — probably only a few belong here.

1 means he did something, but less than the average.

2 means he did an average amount — perhaps nearly half the children belong here.

3 means he did more than the average.

4 means he did a tremendous lot. *Only a few* belong here.

X means you don't know at all what he did.

Please add the names of pupils who are in your class but not on the list. *But mark only for the enterprises during which you had these children.*

As soon as you have marked these pupils for the enterprises that went on while you had them in charge, hand this blank to the teacher of the grade next below yours so that other teachers can mark for the rest of the enterprises.

Results of the Service-Record Technique. Most of the teachers who were asked to report the relative amount of work done by their pupils in social service projects turned in usable records. In a few cases there was no discrimination among the pupils, and in a few others the opportunities were too few to afford any adequate basis for a score which included frequency as well as amount of service. There were also complicating elements in the situation, such as the influence of the teacher and the social pressure of the other children, desire for recognition, competition among classes, and the like. But these factors appear also in any test situation and in real life as well.

In the case of the private school (P), the following projects were represented among the fifth- and sixth-grade classes: collecting books for Hampton Institute, raising money for a local day nursery, collecting clothing for families at Christmas, packing Thanksgiving baskets, raising money for a much-loved helper in the school building.

In the suburban community (A), these enterprises appeared:

providing Thanksgiving dinners, giving to the local hospital, preparing Red Cross boxes, giving to Near East Relief, and several others.*

* No direct reliability figures on these facts could be secured, but the intercorrelations among the scores representing the amount done for these various good causes will throw some light on their significance as measures of the service tendency. These are reported in Tables III and IV for the two communities.

TABLE III

INTERCORRELATION OF HELPFUL ACTS IN A SUBURBAN COMMUNITY (A)

	2	3	4
1. Thanksgiving dinner	.443	.207	.262
2. Red Cross boxes		.214	.095
3. Near East Relief			.306
4. Local hospital			

TABLE IV

INTERCORRELATION OF HELPFUL ACTS IN A PRIVATE SCHOOL (P)

	2	3	4	5
1. Thanksgiving baskets	.685	.418	.485	.727
2. Clothing at Christmas		.586	.531	.656
3. Money for day nursery			.579	.616
4. Book drive for Hampton				.592
5. Gift to helper				

The r's average .254 for Table III and .587 for Table IV. These may be substituted in Formulas (8) and (13) of *Studies in Deceit*, which are as follows, where T = the number of tests and r = their average intercorrelation:

$$r(a_1 + b_1 + c_1 \cdots n_1)(a_2 + b_2 + c_2 \cdots n_2) = \frac{Tr}{1 + (T - 1)r}$$

Formula (8), Spearman-Brown

$$r(a + b + c \cdots n)(s + t + u \cdots \infty) = \sqrt{\frac{Tr}{1 + (T - 1)r}}$$

Formula (13)

Formula (8) gives a predicted reliability of .577 for the suburban scores and

SELF–SACRIFICE TECHNIQUES

The techniques just described were too awkward and slow to justify extensive use; so we adopted two simpler plans, both involving the giving up of a cherished object for some charity. These we tried out in an institution for homeless children (population D).

Giving Up Money

We learned that the children in the institution frequently received presents of money from interested persons. Consequently we arranged with the principal of the school to announce to the children that they were to receive a gift of money from some interested people in New York and that this would probably amount to something like twenty-five cents apiece. He told the children that he would like to know what they were going to do with the money and that he had prepared some cards on which they could state their plans. He also mentioned that they might want to give some of the money to charity, to keep some, and to put some in the school bank. So these three objectives were stated on the card thus:

Name . . . . . .
Self . . . . . .
Bank . . . . . .
Charity . . . . . .

The cards were distributed and each child wrote his name and the amounts he intended to spend on each of the three objects: himself, his bank account, and charity.

When the money arrived, there was nineteen cents for each child instead of twenty-five, nine pennies and one dime, all put up in envelopes for distribution. Before giving out the money, the principal told the children that there would be only nineteen cents instead of twenty-five and suggested that they might like to revise

---

of .877 for the private school scores. Formula (13) gives theoretical validities of .760 and .936 for the two sets of scores. The latter are reasonably satisfactory and suggest the possibilities of this procedure as a valid method of securing facts of this sort.

their statement of how they wanted to spend it. Duplicate cards were then given out and each child indicated how he wished to spend the nineteen cents.

After the second choices had been made, the money was given to the children and they were asked to take care of it. If they wished to bank any of it, the school bank would be open the rest of the day. And in the principal's office would be found a box where they could put whatever they wanted to give to charity. They were asked to write their name on an envelope provided for those who wanted to give away any or all of the money, together with the name of the charity they wanted to give the money to.

Results of the Money Technique. A comparison of what was done on the three occasions — the first vote, the second vote, and the actual giving — is presented in Figures 2 to 6.

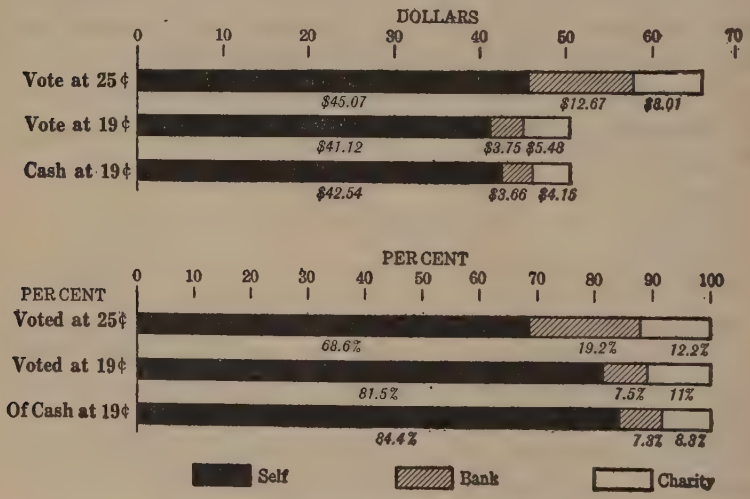


FIGURE 2

Disposition of Money Gift by Vote and Cash

Figure 2 shows the amounts and percentages allotted to self, to the bank, and to charity on the two votes and in actual giving. The per cent of nineteen cents voted for charity is only 1% less

than the per cent of twenty-five cents voted for charity, although the total drop in the amount received is 24%.

Figure 3 shows how this drop of 24% is distributed by the pupils. It can clearly be seen that the bank is the chief sufferer, with a loss

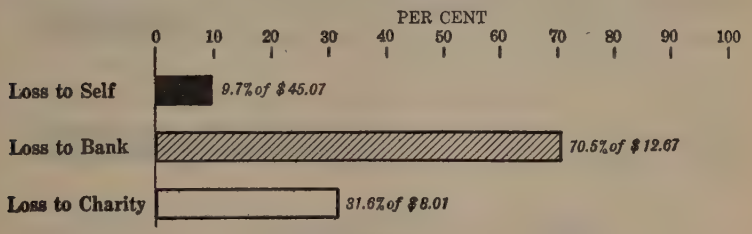


FIGURE 3

PERCENTAGES TAKEN FROM AMOUNTS VOTED TO SELF, BANK, AND CHARITY WHEN GIFT DROPPED FROM 25¢ TO 19¢

of 70.5%, charity next, losing 31.6%, and the amount voted to self dropping only 9.7%.

Figure 4 gives a somewhat different picture in terms of the proportion of the population voting to give all to self, part to self and part to charity, and all to charity, on the two occasions. There is

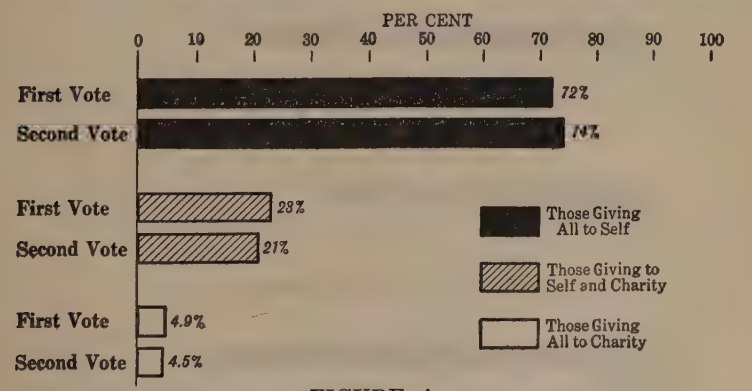


FIGURE 4

PROPORTION GIVING ALL TO SELF, ALL TO CHARITY, AND PART TO BOTH SELF AND CHARITY

N = 266

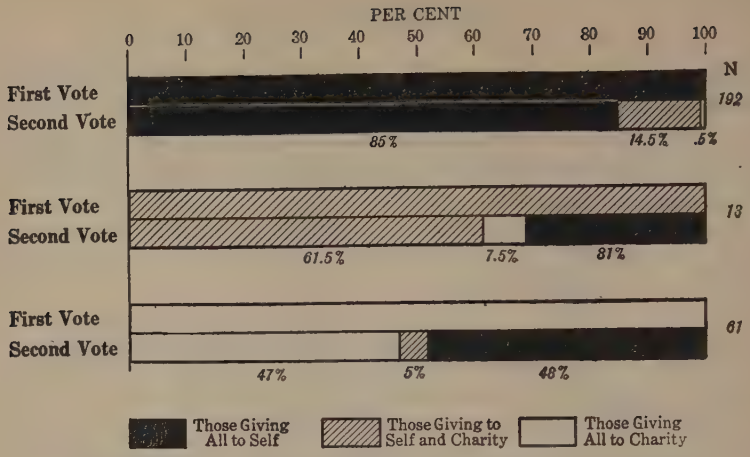


FIGURE 5. Changeability of Givers and Non-Givers

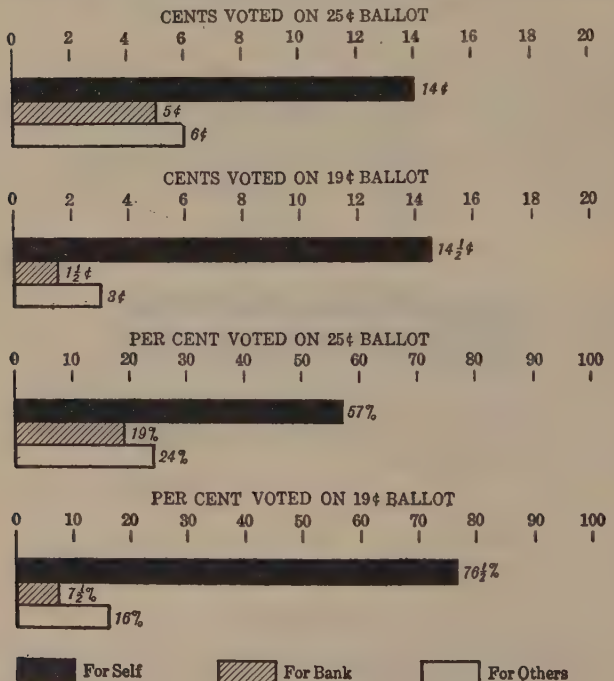


FIGURE 6. Disposition of Money Gift on First and Second Votes

not much change in the numbers giving in spite of the 24% reduction in the gift from twenty-five to nineteen cents.

Finally, Figure 5 shows the relative changeability of the givers and non-givers. Of those voting all to self on the first occasion, 85% did not change on the second occasion, 14.5% divided their nineteen cents between self and charity, and .5% decided to give all to charity. But of those who at first voted all to charity, only 61.5% persisted on the second vote, 31% going to the other extreme and only 7.5% partitioning the amount. Of those who at first divided the gift, 47% continued to do so, 5% voted all to charity, and 48% voted all to self. This tendency to shift to self is more strikingly seen if, instead of counting as unselfish what was voted for " charity " only, we include what was voted for anything not purely selfish, such as Christmas gifts for others or cottage fund for radio. Such gifts could not be traced, of course, but the contrast between the first vote and the second is shown in Figure 6. Not only did the proportion kept for self rise from 57% on the first ballot to 76.5% on the second ballot, but this proportionately larger figure was obtained chiefly at the sacrifice of the bank account. Combining bank and self, we find that on the first vote 19¢ was reserved for self and 6¢ was voted for others, whereas when 19¢ was the total amount available, 3¢ on the average was taken from each category, leaving 16¢ for self and 3¢ for others.*

* The reliabilities of these techniques may be judged from the correlations of Table V.

TABLE V

Correlations between Actual Gifts and Amounts Voted to Self or Charity and to Bank

	Bank and Self (or Charity)	Bank Only
First and second votes	.603	.381
Second vote and cash	.906	.281
First and second votes, with charity broadly interpreted	.587	.381

The r of .906 is somewhat spurious, owing to the fact that some 74% chose

We shall show later that on another test also the activity for self increased and the activity for others decreased when the situation recurred.

Giving Up Ice Cream

The second procedure made use of another experience of the school. On rare occasions the children had ice cream for dessert. Just preceding one such occasion (provided for by us) the principal made a carefully prepared statement to the pupils at the morning assembly. The statement read:

> Boys and Girls: A few days ago a friend came to see me and Dr. M. and said he wanted to do something for you children. Dr. M. suggested that he treat you to ice cream. At this I made a wry face and said: "I think Dr. M. spoils you children by allowing you so much ice cream and sweets. Let's have the money and use it as we see fit." But as usual Dr. M. had his own way and the friend agreed to give you a treat of ice cream.
>
> Now each of you is entitled to a plate of ice cream. Please understand that. But I had a purpose in asking you to bring your pencils this morning or I wouldn't have asked you. First write your name on the slip of paper that is being passed out. Then write "Yes" or "No" to the question I am going to ask you. How many of you would be willing to give up your ice cream for a charitable cause? If you are willing to give it up, write "Yes"; if not, write "No." It doesn't make a particle of difference to me which you write. I shan't even keep track of it. This ice cream costs about five or six cents per plate. Understand now. If you want to give it up and let the money go to a charitable cause, write "Yes"; if not, write "No." It's your ice cream; so do as you please.

---

to keep the money and of these 85% did not change their minds. Considering only those who gave or said they would give something, the r between the statements and the gifts is .864.

Bank deposits were made largely by children who did not plan to do so. Of the twenty-five saying they were going to put some money in the school bank, only nine actually made a deposit that day. Of the $5.48 voted for charities to be given in the way specified by the principal, $4.15 or 76% actually was turned in.

The first ballot was taken. Then the principal said:

Children, every time I try to run off a ballot of this kind I have to do it twice or more before I get it right. I expected this to-day and provided an extra set of slips. I notice there are one or two slips here without names on them. There may be others; so I think we should do it over and get it right. Now please everybody put your name on the first thing you do. It occurs to me that I should tell you to what charity this money is going. If you have read the papers, and you all get the Sunday paper, you may have noticed that there is a fund being collected for the suffering Jewish children in Russia. I propose to send what money we collect to that fund. These Jewish children in Russia are starving. They do not have even bread, much less ice cream. Probably some day we shall receive a letter from these children stating their appreciation. Will the teachers pass out the slips?

After the second ballot, the principal said:

There is always a joy in giving. You feel happier when you have given something. I know that those of you who have given must feel happy about it. I want to thank those who have so generously given up their ice cream that the starving children of Russia might have bread.

If there are any here who would like to change their ballots, they may do so by coming to my office any time to-day before twelve o'clock.

We furnished the ice cream for the last remaining nine who decided they wanted it anyway, and of course the money voted for charity was sent as directed.

Results of the Ice Cream Technique. The value of this experiment consists chiefly in the scaling of the appeal. The first statement made to the children was a matter-of-fact announcement offering them the opportunity of giving to charity the cost of the ice cream they were about to receive. More than half (53.5%) of these institutional children, for whom ice cream was an unusual treat, signed away their dessert. On the second appeal, which stated the object to which the money would be sent, 35.6% came in, making 89%; and after the final statement, in which the principal spoke of the joy of giving and the gratitude of the Russian children, 5% more said they would do without their cream, bringing the total to 94%.

VALIDATION OF THE SELF-SACRIFICE TECHNIQUES

We attempted to secure ratings on our institutional children as a criterion for our tests of helpful behavior. Both the classroom teachers and the house mothers in charge of the cottages where the children lived gave us ratings on the following plan:

The pupils of one group are all listed on the accompanying page. Then follow six columns, numbered in the same way as the questions below on this page. The rater reads question one, then turns to the list and makes the appropriate mark in column 1, right down the list for each pupil. Then question two is read, and answered in column 2 for each pupil, etc.

The mark that is to be made is simply a letter, indicating the frequency with which a child does the thing asked about in the question. The letters are the initials of the words *Always*, *Usually*, *Sometimes*, *Rarely*, *Never* — A, U, S, R, N. The rater puts one of these five letters opposite each name and in each column, unless he does not know the facts. In this case he puts an X in the column instead.

The marks for an imaginary character are entered as an illustration at the head of the list. Do the same with the rest. Do not be afraid to use the letters A and N if the facts justify it.

The questions to be answered are as follows:

1. When money is being raised for anything, does he give any?
2. Does he give till it "hurts"?
3. Does he help others with their lessons?
4. Does he help with cottage duties without expecting anything in return?
5. Does he go out of his way to be nice to pupils whom he doesn't like?
6. Does he share with others things that he could just as well keep for himself?

In spite of the unusual opportunity these raters had of observing the children, the correlation between their ratings and helpful behavior was $-.034$. This indicates either that the test behavior was by no means typical of what these children usually did or that the raters did not know them very well.*

* Theoretical validation was computed, as usual, by the use of the formulas quoted in the footnote on page 28. The basic inter-*r*'s are given in Table VI.

SCALING SELF-SACRIFICE

In Chapter XXI of Book One and Chapter XII of Book Two, *Studies in Deceit*, we had occasion to describe the way in which the attitude toward a certain type of deceptive act tends to scale. It was found that those whose deception tendency was strong enough to overcome greater resistances to deception also overcame lesser resistances, much as a bullet of a velocity sufficient to penetrate a piece of metal will also penetrate a piece of wood of the same thickness. In other words, cheating attitudes seem to be specific and as much functions of the situations as are the overt acts.

In the case of these self-sacrifice techniques the attitudes seem somewhat more general. The two types of behavior, giving up money and giving up ice cream, which correlate only about .15, fall into a common scale. This is possibly due to the relatively simple pattern of the helpfulness involved. Cheating attitudes are complex, involving various ulterior motivations and various conflicting codes. They operate, furthermore, against social disapproval and supervision. Helpfulness of the sort reported, on the other hand, may be less dependent on ulterior motives and more directly related to its object as its end. In our case, also, it had the support and approval of the community. Like cheating, it was a positive act. It did something to someone. Furthermore, in each case, self-denial was involved.

TABLE VI

INTERCORRELATIONS OF SELF-SACRIFICE TESTS

	Second Vote	Cash Gift	Ice Cream
First vote	.603	.617	.176
Second vote		.906	.137
Cash gift			.127

These r's average .428. Substituting this in the formulas, we find a predicted reliability of .747 for the general procedure and a predicted validity of .865. Omitting the ice cream scores and considering only the money gifts, the reliability is .879 and validity .937.

Study of the percentages exhibiting helpful behavior at the various stages of the money and ice cream experiments resulted in the combinations pictured in Figure 7, which is comparable to Figure 95 of *Studies in Deceit.*

Certain of the spaces are filled in by logical necessity. Those who gave all their money of course gave also a part, and those voting away their ice cream on the first appeal naturally were included as the appeal was increased.

Counting these logically filled spaces as realities, the scale is 91% complete.*

THE RELATION OF PRELIMINARY MEASURES OF HELPFUL BEHAVIOR TO MISCELLANEOUS FACTS

In anticipation of the results to be reported from our later techniques, it may be of interest to see the general trend of association between altruistic behavior and a group of social and biological facts. In the private school and the institution there is a slight tendency for the older children to be more coöperative. In the

* Details of the procedure for scaling the values of the steps are given in Appendix A.

This process resulted in a scaled helpfulness score representing theoretically the attitude of the subjects toward the situations represented in the test. Using these scores as a criterion for the separate tests, we find the *r*'s of Table VII.

TABLE VII

CORRELATIONS BETWEEN SELF-SACRIFICE SCORES AND SCALED SCORES

	r
First vote of money	.594
Second vote of money	.654
Cash gift	.645
Ice cream score	.708

These *r*'s justify us in assuming that these tests are measures of the same tendency and so may be regarded as equivalent parts of one technique, as assumed in the footnote on pages 36 and 37.

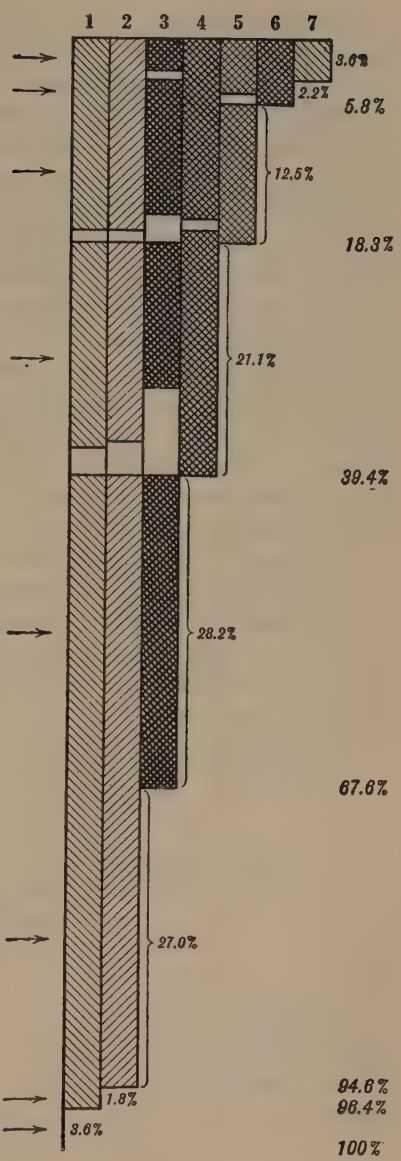


FIGURE 7

The Scaling of Helpful Behavior

The numbers at the top refer to the following test results, which are described in Chapter I:

1 Giving up ice cream — third appeal
2 Giving up ice cream — second appeal
3 Giving up ice cream — first appeal
4 Voting to give away some money
5 Giving away some money
6 Voting to give away all the money
7 Giving away all the money

The amount of resistance at each step is shown by the density of the shading. 3.6% of the cases overcame all the resistances. 3.6% overcame none and did nothing at all, as shown at the bottom.

The arrows show the direction in which the charity drive is operating. The figures at the right give the percentages of the whole group passing through each level of resistance, and the blank spaces show how much these percentage figures should be reduced to indicate the exact extent to which this type of behavior actually scaled. The total blank space is only 9% of the chart — *i.e.*, the scale is 91% complete. For further explanation, see Appendix A.

suburban community, on the other hand, the tendency is the other way, with about a year and a half difference in age between the 10% who were most generous and the 10% who were least generous.

There is still less association between intelligence and helpfulness, though the more intelligent tend to be slightly the more helpful.

As for sex, in the institution containing 143 boys and 78 girls, only 28% of the boys reached or exceeded the median helpfulness score of the girls — or to put it the other way around, about three-fourths of the boys were less helpful than the average girl. In the suburban community there was less difference, 44% of the boys reaching or exceeding the median scores of the girls; and in the private school there were no sex differences.

No association was found between helpfulness and physical condition. The more suggestible children, as measured by the Otis Test,* were also slightly less helpful; and to about the same degree higher home background, as measured by the Sims Score Sheet,* was associated with greater helpfulness.

There was no resemblance between brothers and sisters in the tendency to be of help, which might be interpreted as meaning that the homes were of small influence in the matter.

School attendance was also apparently of no moment, as the correlation between years in the private school and amount of helpfulness was .092.†

* Described in detail in *Studies in Deceit.*

† All these associations are represented in the *r*'s of Table VIII.

TABLE VIII

CORRELATIONS BETWEEN SELF-SACRIFICE SCORES AND VARIOUS FACTS

	INSTITUTION (D)	SUBURB (A)	PRIVATE SCHOOL (P)
Age	+ .237	− .315	+ .141
IQ	+ .110	+ .195	+ .251
Physical condition	− .096		
Resistance to suggestion (Otis)		+ .274	
Home background (Sims)			+ .257
Years in one school			+ .092
Siblings	+ .083		

When the various helpfulness scores are made equivalent, in terms of the average amount of help given to all the opportunities, we find the institution children rating lowest, with a score of 1.1, the suburban children next, with a score of 1.4, and the private school highest, with a score of 2.1. The differences between the schools are statistically reliable. Nevertheless, the institutional children were most responsive in certain situations. Following a letter of thanks from the charity to which their ice cream money was sent, these children decided on their own initiative that the few dollars they had contributed were not sufficient; and they set out to raise from their slender resources the amazing sum of $110 as their quota and then went over the top to a total of $152.25. This illustrates the difficulty of setting up any test situation as an adequate sample of even a single type of behavior.

CHAPTER II

ANALYSIS OF SERVICE AS AN OBJECT OF STUDY

BEHAVIOR SCALES ORIGINATING IN CONFLICT

We had occasion to point out in Volume I that deception seemed to arise in situations involving conflict. The more deep-seated the conflict, the greater the amount of deception. This analysis was borne out by the chapters dealing with the relation between school spirit and honesty. Situations characterized by an atmosphere of friendliness and coöperation were found to be far less provocative of deception than those involving antagonism between pupils and teachers.

In recent years students of human nature have been emphasizing the significance of those motives which lead to conflict. One does not need to go back to Hobbes for the theory that self-interest and the desire for power and possession are the dominant forces in social and individual life. Largely under the influence of the Freudians, the effect on personality which follows the thwarting of these fundamental drives is increasingly recognized,* and the social consequences of their unchecked activity have been vividly portrayed by Bertrand Russell,† Will Irwin,‡ and others. In resolution of this dilemma, evidence is accumulating on the one hand that self-interest is not only not the ultimate motive, but is not even the strongest motive,§ and on the other, that conflict,

* Cf. Thomas, W. I., Chapter VI of *The Unconscious*, a symposium edited by Mrs. W. F. Dummer.

† *Why Men Fight.* ‡ *The Next War.*

§ Kropotkin, P., *Mutual Aid; a Factor of Evolution.* W. Heinemann, London, 1919.

once recognized, may give way to coöperation as a more satisfactory technique for achieving the ends of life.*

In Book One of this volume, we are studying these fundamental characteristics of the situations to which deception and honesty seem to be related. We have already referred to these social facts as " coöperation," " service," " altruism," " helpfulness," and so on. A more detailed analysis of what these terms represent would seem to be desirable at this point.

If we begin once more with " conflict " as a limiting term, we can follow out two lines of development toward an ideal social situation. One of these moves from the completely selfish behavior of a state of conflict, in which every man's hand is raised against every other, to one in which the individual freely gives all to the service of humanity, even life if necessary. We shall call this other end of the scale " charity," in the New Testament sense — again a limiting term, and implying no thought of selfish interest whatsoever. In the other direction we may move toward a state of affairs characterized by free and complete coöperation. On the charity scale we move from " self " to " others," with the emphasis changing from " against " to " for." On the coöperation scale we also move from self to others, but with the emphasis changing from " alone " to " with."

In every area of experience, the steps on these imaginary scales would be represented by typical acts or social structures. We have tentatively illustrated one such series in Figure 8. It may be noted that the honesty scale also reaches out from its origin in conflict toward a status of absolute integrity. The honesty scale is therefore included.

On the coöperation scale, which extends upward, we have represented massacre or marauding as one step removed from complete individualistic conflict, where each man is for himself. In a genuine panic, the original situation frequently recurs. There is no organized movement. But in a massacre or a raid there is usually some degree of coöperation among the aggressors. They work together

* "A Coöperative Technique for Conflict," a pamphlet published by The Inquiry, 129 East 52d Street, New York City.

in order to destroy. War, with no fewer than two parties organized, has in the past represented a higher stage of coöperation, for the two groups must at least have agreed to fight. Otherwise, the conflict again became massacre. Modern warfare, with its attacks on non-combatants, has been offering notorious examples of reversion to the level of massacre.

As we ascend the scale we come to a more highly organized state of affairs, where conflict is so surrounded by customs and rules as to

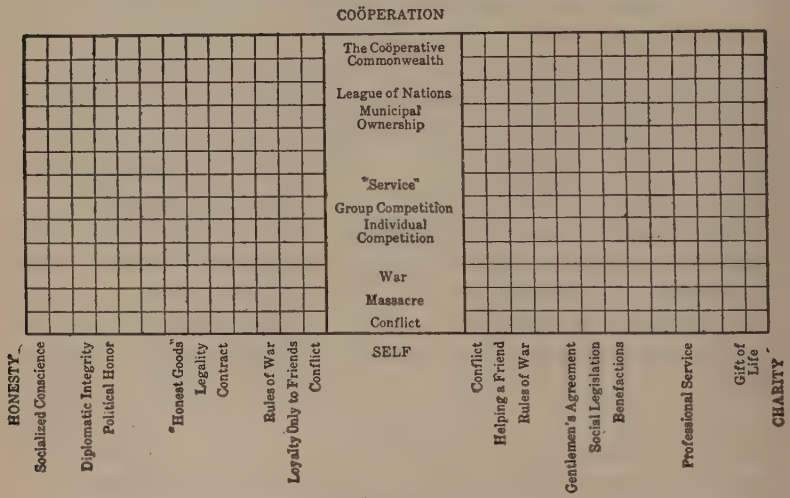


FIGURE 8

Imaginary Behavior Scales with Zero Point at a State of Social Conflict

justify the term " competition." In its lower stages it is again every man for himself, but groups ultimately organize, much as in war, and compete with one another, firm against firm, class against class, industry against industry, nation against nation. This in turn gives way to a kind of coöperative competition, with price agreements and so on, and a certain pride in " service." There is no illusion anywhere as to motive, but the day of " the public be damned " is beneath this level.

Still higher up we come to operations involving no direct conflict at

all, save as communities may compete with one another for trade and industry by raising their own standard of living through such coöperative enterprises as municipal lighting, water, power, and sewerage.

Among nations this stage might be represented by the League of Nations, with a still higher level, from which conflict is completely eliminated, in one coöperative commonwealth of free men.

On the charity scale, the first step above conflict with all men in the interest of self is the willingness to help a friend. The massacring hordes assist one another to murder their enemies. In war, however, there are certain restraints not recognized in massacre. By the rules of warfare, the enemy has certain rights. Prisoners are reasonably cared for. The enemy sick are given medical aid. In a fashion, the grim humor of which is lost upon the military mind, the contending sides eagerly accommodate one another in offering themselves for target practice.

Above the fantastic level of war, we find increasing dependence on rules limiting the behavior of competitors. Before these have reached the stage of law they are often formulated in gentlemen's agreements. Beyond legal requirements, in the way of generosity, are the voluntary benefactions of the wealthy, which, while not leaving them poor and thus involving no great personal sacrifice, nevertheless must be counted as in the direction of charity.

Still higher on the scale is the noncompetitive profession of paid social service with all its branches. To be sure, among professional groups which depend on fees, there is still individual competition, but its rigors are somewhat mitigated by law and code. At the top of the list should be placed the missionary, who, on a meager living, gives his entire time to others; and from this group, as well as from many other walks of life, are drawn those who, with no return at all, give energy, time, property, and often life itself.

In the third direction we come first to such mild loyalty among friends as gives rise to the myth of honor among thieves. Honesty at this level means loyalty to the group. Next come the rules of war, with the corresponding possibility of treachery. Higher up, on the level of competition, we have first the contract and its fulfillment or violation; then general legislation, with observance

and non-observance; then, beyond the requirement of law, a sense of honor working out in the production of honest goods. Its degradation appears in such a practice as selling imperfect or adulterated articles as "honest goods" — a practice which long ago made "shoddy" a term of reproach.* Higher still comes political honor and its opposite, corruption, and higher still, diplomatic integrity, and finally the socialized conscience of mankind.

Generally speaking, there is a high correlation among these levels of social growth, although, above the level of conflict, one would find individual exceptions to this rule. One could not have a coöperative commonwealth without a high degree of unselfishness and honor; but mixed in with war one finds heroic self-sacrifice, or a fine sense of personal honor in the conduct of a competitive business. But these remain exceptions, and social growth away from conflict advances toward coöperation, good will, and honor in one inclusive movement.

THE SIGNIFICANCE OF MOTIVE

It may be a matter of surprise that so little attention is here paid to the motives that may be prompting behavior of a coöperative or charitable nature. As in the case of our study of deception, we do not regard motive as of less importance but only as more difficult of approach. We begin with the objective facts. Without knowledge of actual tendencies and habits, knowledge of motives would be of small use. Furthermore, it is easier to pass from deeds to their causes than from motives to their possible results. A man may be of great service to society even while working solely for himself. It is equally true that he may be a great nuisance while devoting himself heart and soul to the good of his neighbor. The neighbor may prefer to be let alone. Our sole standard by which we rate any behavior is for the moment its utility in the given situation. We assume that the man of socialized motive is the one who seeks to discover what this utility is and then to produce it.

* Shoddy has since become a legitimate commercial product, recognized for what it is.

There are times, of course, when circumstances intervene between a motive and its appropriate act, either preventing its occurrence or distorting it so completely as to make it necessary for the one responsible to disclaim the intention implied in the act. Thus in passing a glass of water one may be jostled in such a way as to spill it and ruin a gown. Only a crank could see in this any relation to the original intention, which was to help, not to injure. Of course it is obvious that the act *does not* correspond to the motive. From this extreme one moves gradually into areas of behavior in which the distinction is not so clear cut, as when one performs an act of charity which results in pauperizing its object, euphemistically called the beneficiary, or when one introduces into his factory or business a scheme for increasing human efficiency and finds that he has stirred up ill feeling among his employees. In such cases, the fault is not different in kind from that of the spilled water. It differs only in the nature of the intruding circumstance. We spoke of the first mishap as an accident. We might have recognized that greater caution — that is, greater attention to all the attendant circumstances — would have prevented the accident and enabled the well-wisher to realize his first intention. The adequacy of his motive would have been represented in the adequacy of his control over the conditions for its expression.

So in these other cases, the intruding circumstances were the result of ignorance. More careful foresight of all the consequences — that is, greater interest in the result aimed at — would have made possible a more adequate control.

There is, thus, even in theory, no break between an act and its motive. The two are phases of one process of life adjustment.*

THE SIGNIFICANCE OF CONFLICT, COMPETITION, AND COÖPERATION

We have illustrated the movement from conflict to coöperation and from "taking" to "giving" in only a general way. This

* Cf. Dewey, John, *Human Nature and Conduct*, Chapter III; Symonds, Percival M., *The Nature of Conduct*, Chapter XI.

movement does not necessarily begin at the bottom and proceed upward. Gregarious animals such as man do not begin at the bottom, although they may completely lose their coöperative habits in a panic, as has been suggested, or slough off social restraints now and then when some strong self-interest demands satisfaction. In some families the members seem to be in an almost perpetual state of conflict. Some classrooms seem never to have risen above the level of war or possibly of group competition. In view of the fundamental autocracy of our school systems, this is only to be expected. Whenever government is created without the consent of the governed and acts without the rightful participation of the governed, the persons governed are in a state of potential or actual rebellion and civil war is never far out of sight. That the conflict between governed and governing does not break out openly is evidence only of a nice balance of contending forces which may be easily disturbed, as our recent student strikes testify. Under these circumstances, a high level of coöperative behavior cannot be attained.

In marauding gangs, the volunteer membership often develops a closer coöperation than a school classroom and so, with a motive perhaps no more hostile to " society " than that of a recalcitrant schoolroom, may give a more intensive training in citizenship than the school. Similarly, volunteer groups within the class offer opportunities for coöperation at least on the level of group competition. Or when class is pitted against class in some school contest, the room is transformed at once into a coöperating unit, the members of which help one another to overcome the common enemy.

Competition may thus be an instrument in the creation of a coöperative group which operates on a higher level than that of individual self-interested conflict. But few would claim that organized competition, the motive for which is to " destroy " the opposing group, represents coöperation at its best. One must turn to other types of group experience for illustrations of thoroughly socialized activity. We find these in the volunteer clubs, or democratically appointed committees, or self-governing groups, the

members of which work out together the line of action which constitutes their reason for existence and divide up the work of carrying out their purposes in some fashion agreeable to all. The result is happy, free, successful, joint action. When this action is in the interest of a larger group, as when a club works for civic betterment without thought of what the club itself is to get out of the activity, we have coöperation at its best. The integration of such coöperating groups in a world society would constitute the great objective of political effort on this plane.*

To reach the outer ranges of coöperative behavior, we have a long way still to go in school and home, in industry, and in government. Consequently we can hardly expect any measures of these tendencies and abilities to go very high on a scale of coöperative and charitable behavior. In the next chapters we shall show the ranges of the scale we have attempted to include in our tests and the types of behavior we have been able to standardize to some degree of objectivity.

PREVIOUS EFFORTS TO MEASURE CHARITY AND COÖPERATION

We have been able to discover very few precedents in this field of measurement. There are two unpublished devices mentioned in our review of techniques up to and including 1925.†

In one of these, the situation used was a party given to parents in which the children had an opportunity to behave in helpful ways. A record was kept of what each child did. In the other, the pupils agreed to raise lilies for shut-ins. The tendency to bring pleasure to others was presumably measured by the growth

* For a most stimulating discussion of the nature of democratic group experience, see M. P. Follett's *The New State* and *Creative Experience*. Cf. also two articles by G. A. Coe on the "Functions of Children," in *Religious Education*, 1918, **13**, 26–32, and 1921, **16**, 29–36.

† "Objective Methods of Measuring Character," *Pedagogical Seminary*, 1925, **32**, 45–67.

of the lily inasmuch as this depended on the persistence of the child's interest in caring for it. These were acts of the self-charity type, involving no coöperation.

In our first plan for a children's stunt party to consist largely of test situations, we included an opportunity for each child to give up all or half his ice cream for some charity to be determined by the group.* In our preliminary experiments with the party idea, this worked very well, but we could not incorporate it in our general test program on account of the brief time available for our deception parties and the great expense involved in their administration.

In an experiment reported by Müller,† the effect of three motives was measured in terms of the amount of work done. Three tasks were used, threading mazes with a pencil, eyes open, cutting out geometrical forms, and simple additions. The basic motive given was, " Work as fast as you can." The succeeding motives were instilled as follows: (1) the will to excel — the members of the group compete with one another; (2) practical utility — a discussion of the value of the exercises for self-improvement; (3) altruism — *e.g.*, " If you do exceptionally well to-day in additions, I will give you something which you will give to someone as a present." Müller found that under the so-called altruistic motive the improvement was slightly greater than under the so-called competitive motive. She used only 50 cases, however, and it is not clear to what extent the improvement reported was due to practice or, in the case of the altruistic motive, to other factors than the desire to help someone else. It is difficult to tell just how a child would react to the directions, " I will give you something which you will give to someone else as a present." Quite apart from his desire to help someone, he would naturally be eager to find out what the " something " might be.

---

* *Studies in Deceit*, Book Two, p. 266.

† Müller, Johanna, "Versuch über die Einvirkung von Motiven auf Körperliche und geistige Leistungen bei Schulkindern," *Zeitschrift für angewandte Psych.* 1924, **24**, 81–128.

Sims,* using two types of test, shows that improvement in scores is influenced more by the desire to improve one's score, when this is fortified by knowledge of one's progress, than by the desire to help one's team.

Still less has been done to measure the tendency to work either for or with a group. Washburne† reports an interesting experiment in a school in Russia. Each child has his own garden plot but also has a joint interest in a common garden tended by all the children. Records are kept to show the relative amount of interest in the two types of work. No results are given of this ingenious scheme.

McCaskill devised a technique for use in the New York Y.M.C.A. survey by means of which boys who preferred to work for individual prizes could be distinguished from boys who preferred to work for a prize for their own club.‡ At the beginning of three series of tests on knowledge of sports the examiner told the boys about the two prizes, one for the boy who scored highest and one for the club which scored highest. The tests of each series were alike as to difficulty, but test A was to count 250 for the club and 50 for the individual; B, 50 for the club and 250 for the individual; C, 200 and 100; D, 100 and 200; E, 225 and 75; F, 75 and 225; G, 25 and 275; and H, 275 and 25, respectively. Each might select five tests to answer from each series. The score equaled the per cent of the subject's club score of the total possible club score. In practice it was found that there was only 5% disagreement between two of the tests in the loyalty scores derived from each.§ It was found that the clubs more recently organized had lower loyalty scores than those that had been organized for a longer period.

It should be noted that the activity represented in these experi-

---

* Sims, V. M., "The Relative Influence of Two Types of Motivation on Improvement," *Journal of Educational Psychology*, 1928, **19**, 480–484.

† Washburne, Carlton, "The Good and Bad in Russian Education," *New Era*, 1928, **9**, 8–12.

‡ See Swift, A. L., Jr., *The Survey of the Y.M.C.A. of the City of New York*, Association Press, 1927, limited edition.

§ Swift reports an r of .82 between two forms.

ments is genuinely coöperative in character and is measured in relation to activity undertaken in one's own interest. In a sense, the scores represent the distance along the coöperation scale to which an individual will rise above the point where he works entirely for himself. The highest step on the scale which can be reached in such behavior depends on the nature of the group activity. In the case of the Russian experiment, the group action is entirely noncompetitive and is high up on the scale, but in the case of the McCaskill test the level is much lower and, in view of the identification of the individual with the group interest, is in truth a mixture of selfish and unselfish motives.

Several experiments * have shown that rivalry increases the intensity of individual activity over the normal level even when no additional awards or prizes are offered for improvement, and it is also possible that the mere presence of others tends to stimulate individual activity above the norm of solitary work, whether these others be participants or spectators.†

Consequently when group is pitted against group, even though the individuals are not consciously working to improve their own scores, there is presumably an effect from the competitive aspect of the situation which accounts in part at least for whatever coöperation takes place among the members of each competing group.

Our own techniques are of two types, corresponding to the two sorts of social development above the level of conflict. We shall not attempt any hard and fast classification, however, but shall devote the next few chapters to a description of our procedures and certain preliminary indications of their value.

---

* See Allport, F. H., *Social Psychology*, Chapter XI; Hurlock, Elizabeth B., "The Use of Group Rivalry as an Incentive," *Journal of Abnormal and Social Psychology*, 1927, **22**, 278–290; Müller, J., *op. cit.*

† Allport, F. H., *loc. cit.;* also Sengupta, N. N., and Sinha, C. P. N., "Mental Work in Isolation and in Group," *Indian Journal of Psychology*, 1926, **1**, 106–110. See *Psychological Abstracts*, 1927, **1**, 73.

CHAPTER III

THE CEI SERVICE TESTS

We have described in the preceding chapter the two types of helpful behavior which may be scaled upward from an absolute zero point of purely selfish conflict. On the one hand there is the variable tendency to do things for others rather than for oneself, and on the other hand there is the variable tendency to work with others rather than to stand alone. This second way of acting passes into and through a stage of *coöperation for the sake of organized competition* to a higher level of *coöperation for a noncompetitive object*, the significance of which lies in the relation of the coöperating individuals to one another. Coöperation and charity, as thus defined, are usually but not always mingled. One sometimes does things *for* others without considering what the others really want done. And again, a person may work *with* others in order to accomplish some purpose which he shares with them without intending to be of any direct assistance to the other members of the group. Neither mode of behavior taken by itself is as socially useful as the combination of the two.

In our attempts to devise test situations, we have had these two modes of response in mind. In some tests it would be possible for a child to work on the charity motive solely — the situation calls for no direct coöperation. In none of our coöperation tests, however, is the charity aspect of the behavior absent, for the pupil chooses whether he will help his classmates or work for himself. We did not succeed in building a test which would reach the higher levels of coöperation, but had to content ourselves with such coöperation as was called into play through an interclass contest. In this respect, our toy-making and picture-pasting techniques had features of value which are missing from these

later techniques, but even here there was no actual working together on a common project: each child made his own toy. We were unable to develop a completely coöperative situation in which each child could have a share equivalent to that of every other child except by a rotation of responsibilities within the group. Such rotation could be handled only arbitrarily, however; and furthermore, the time involved was prohibitive for practical uses. What is needed is a technique involving the coöperation of the members of a group in a project to help someone else. The project should consist of small units of work requiring the members of the group to work with one another, not independently. These contributing elements of the project should be voluntarily rotated among the members so that each will have a chance to do each part of the work. These parts should be susceptible of quantitative measurement in terms of the product and in terms of the coöperative skill actually shown during the process of producing the objects required for the project.

Our first set of tests was tried out in grades 5 and 6 of population P. We were greatly aided in this preliminary experiment by the generous coöperation of the teachers, whose intimate knowledge of the pupils was invaluable as a criterion with which to compare our test results. A large number of parents also coöperated with us in supplying information concerning the helpfulness of these children at home.

Each class was treated as a unit and the procedure used was identical in all six groups. The battery * contained the following items, used in the order named:

The Self-or-Class Test
The Money Voting Test
The Learning Exercises
The School Kit Test
The Envelopes Test

* In addition, there was a paper and pencil test, which listed fifty situations with four possible ways of responding to each one. From these imaginary responses the subjects were to choose the one which was most useful, sensible, and helpful under the circumstances. As this was a test of comprehension rather than of behavior, it will be dealt with in the third volume of the report.

THE SELF-OR-CLASS TEST*

A spelling contest was set up in which two sets of prizes were offered, one for the winning classes and one for the winning individuals. No pupil could enter both contests. Each had to choose whether his score was to count for himself and help himself toward getting a prize or count for the class and help the class get a prize. The teachers administered the tests as follows:

The school has had an offer of several prizes. First there are three prizes of $5, $3, and $1, to go to the classroom making the highest score in spelling. This is a classroom prize. Then there are three individual prizes of $3, $2, and $1, to go to the individuals making the highest scores in spelling. It will not make any difference how good or how poor a person or a class is to start with. Only the relative score will be counted and each child or each class will receive a handicap according to ability. No one can enter both contests. Each child must choose whether he will help his *class* by taking the test or help *himself* by taking the test. As both prizes are being offered you are perfectly free to choose.

Now each one must choose for himself; so I have little slips of paper like ballots on which you may say which contest you are entering, for the class prize or for the individual prize. (Pass the slips.) Write your name on the slip and put a check in the first square if you want to help your *class* win a prize and have your score count for the *class*, and put a check in the second square if you want your score to count for an *individual* prize on the chance that *you* may be successful and win the $3 or $2 or $1 for yourself. Your score will count for the class or for yourself, not for both. Is that clear?

The slips were as follows:

Name_____

Grade_____

Check which contest you wish to enter.

(1) ☐ I wish to help the class win a prize.

(2) ☐ I wish to compete for an individual prize.

* Suggested by the McCaskill device described in Chapter II.

Form 2 of the Morrison-McCall Spelling Test of fifty words was used. Allowance was made for age and grade in scoring. The prizes were awarded as indicated.

This technique, which was later displaced by the Free Choice test, would be improved by announcing that the class score will depend in part on the number coöperating. This would prevent anyone choosing the individual contest because he is a poor speller and does not want to hurt the class, as happened in the case of some of the pupils of school P.

THE MONEY VOTING TEST

In the school P experiment this was used twice, first with the winning classes in the spelling contest and second with all the competing classes in the Learning Exercises, to be described presently. On the first occasion, the prize money was in the teacher's hands when the children voted what they wished to do with it, but on the second occasion the situation was more hypothetical. After the Learning Exercises were finished, the teacher suggested the possibility of the class winning a prize and proposed a straw vote, in advance of any discussion, to see how the class would dispose of the money if it should be won. The ballot was as follows:

Write your name on the slip and then mark 1 what you would most like to do with the money. Mark 2 your second choice. Mark 3 your third choice. Mark 4 your fourth choice. Mark 5 your fifth choice.

Name_____

[4]_____ Give all the money to the boy or girl scoring highest in the test.
[2]_____ Buy something for our school, such as bats, balls, skipping ropes, a big picture.
[3]_____ Buy something for the room, such as a picture, a globe of goldfish, some plants.
[5]_____ Divide the money equally among the members of the class.
[1]_____ Buy something for some hospital child or some family needing help or for some other philanthropy.

The following wording was finally adopted for the introduction

of the vote, which constituted, as can readily be seen, a test of generosity or charitableness:

> Do you think our class will get a prize? We might. If we do, what shall we do with the money? Let's take a vote now just to see what everyone thinks he would like the class to do with the money in case we win a prize. I have some slips of paper here showing ways of spending the money.

This was the form used in all subsequent testing, and the teacher of the classroom administered the test, confining herself to these words and permitting no discussion until the ballots were collected. Afterwards, of course, the winning classes were free to discuss what they would do and change their vote if they saw fit.

Scoring the Money Voting Test. A theoretically correct arrangement of the items on the ballot was determined partly by reference to the way the children actually voted and partly by the judgment of the staff of the Inquiry as to the relative social significance of each choice. This "correct" method of voting or ranking is indicated in brackets on the copy of the ballot printed on page 56. If a pupil, in voting his choice of alternatives, gave an item the rank shown in the key, this item was scored 2. If it was ranked one place out of position, it was sometimes dropped to a value of 1 and sometimes the value of 2 was still given. But if it was more than two steps removed from its correct rank, it was usually marked 0. This gave the various scoring values shown in Table IX.

The key on the ballot shows that the fifth alternative, to buy something for someone in need, was rated highest. Hence if the subject rated it 1, he received a score of 2. The table shows that if he made this his second choice he received a score of 1. The second item in the ballot happens to have second place in the key. So making this second choice gives a score of 2, and the table shows that putting it in either first or third position gives a score of 1. But if it is voted fourth or fifth choice, the score is 0. The scores thus range from 0 to 10.

This test gave us very interesting results and was retained in our final battery.

EXERCISE 1

Put under each design the number which that design has in the key. Work across the page. Do each one as you come to it. Wait for the signal to begin.

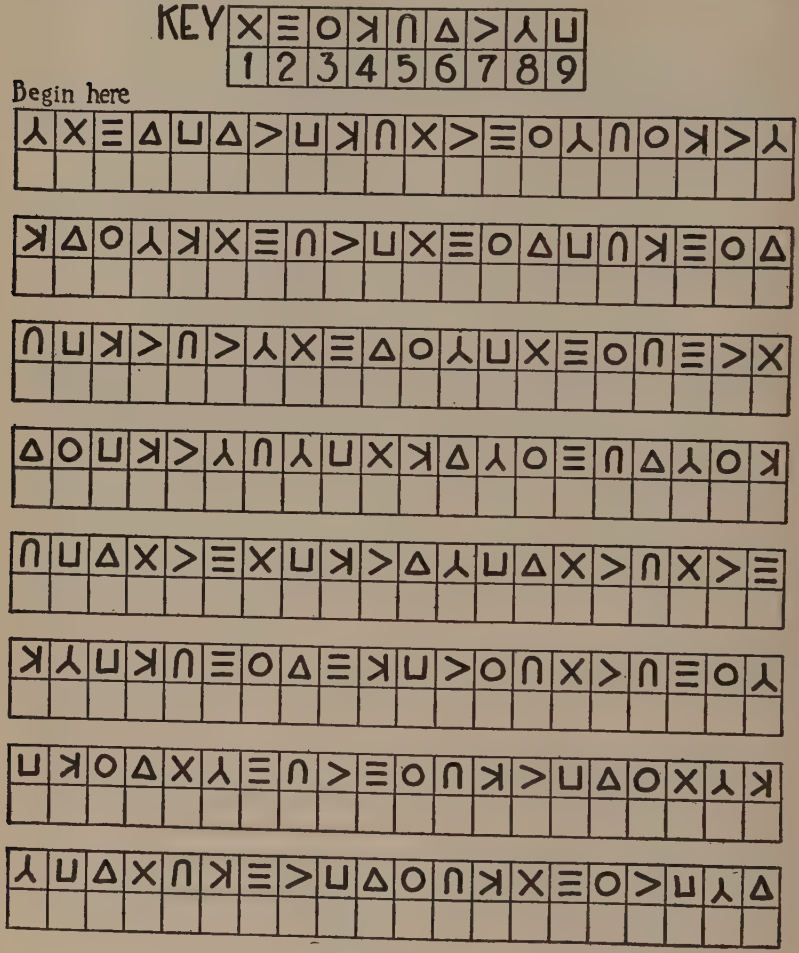


FIGURE 9

ILLUSTRATION OF LEARNING EXERCISES

TABLE IX

SCORING TABLE FOR MONEY VOTE *

ALTERNATIVES	RANKING GIVEN BY PUPILS				
	1	2	3	4	5
First item	0	0	0	2	2
Second item	1	2	1	0	0
Third item	1	1	2	0	0
Fourth item	0	0	0	2	2
Fifth item	2	1	0	0	0

* The score values of the Money Vote could, of course, be scaled more minutely by the usual psycho-physical techniques.

THE LEARNING EXERCISES

This test (used only in population P) depended upon the varying effects of different motives on speed of learning or bond forming. The digit-symbol matching technique as used in the National Intelligence Test was adopted. Figure 9 will illustrate the material. Four forms were prepared as nearly alike in difficulty as could be predicted without actual trial. The relative difficulty was roughly determined by giving all forms to four classes of pupils of about the same age.

Each form of the Learning Exercises had two parts consisting of three pages of symbols. In administering the test, two minutes were allowed for each page or exercise. On the first day no motive was given and the pupils were told that the scores did not count on their grades. On the second day a social motive, helping the Red Cross, was introduced. On the third day the pupils had a chance to help their class win a prize, and on the last day each had a chance to earn a quarter and win a prize, according to the amount of improvement he showed. The essential parts of the directions for each day were as follows:

First Day — Form A, Parts 1 and 2

Examiner says: "We are to have some exercises in learning to-day. Do not do anything with the booklets until I explain what to do." (Part 1 folders are passed.)

Examiner says: "These exercises do not count on your grades. I do not expect you to think that they are either hard or interesting. All I ask is that you start working when I say, 'Go,' and stop immediately when I say, 'Stop.' Pencils up, like this."

The motive is repeated for Part 2.

Second Day — Form B, Parts 1 and 2

Examiner says: "We are to have some more learning exercises to-day, similar to those we had yesterday. Do not do anything to the booklets until I tell you to." (Part 1 folders are passed.)

Examiner says: "These exercises do not count on your grades. But perhaps some of you would like to do a good turn. I'll give you a chance. I have some quarters here which I will let you send to the Red Cross if you will earn them. You can earn them for the Red Cross by making a better score on this work. For every pupil who makes a genuine improvement to-day, I'll send 25 cents to the Red Cross and this class may have the receipt as soon as I can get it. If you want to help the Red Cross, try very hard to do your best on these next trials. Do not open the folders until I give the word. Remember to keep your pencils up when you are not working and to start when I say, 'Go,' and stop when I say, 'Stop.'"

The motive is repeated for Part 2.

Third Day — Form C, Parts 1 and 2

Examiner says: "All the papers are being scored as rapidly as possible. We have a third set of exercises for to-day. Do not do anything to the booklets until I tell you to." (Part 1 folders are passed.)

Examiner says: "I expect these exercises are getting tiresome. Of course they do not count on your grades. But perhaps you would like to help your class win a prize. There is to be an interclass contest to see which class can make the most improvement in its score. The first prize will be $5 and the second prize $3 and the third prize $1. Everyone can improve his score, no matter how good or how poor he is. The winning classes can spend the prizes just as they wish to — for some philanthropy or for something for the class or school. If you want your class to win, work hard to-day. Do not open the folders until I give the word. Pencils up."

The motive is repeated for Part 2.

Fourth Day — Form D, Parts 1 and 2

Examiner says: "Here we are again. Don't be discouraged. This is the last exercise. Do not do anything to the booklets until I tell you to." (Part 1 folders are passed.)

Examiner says: "Now you have been working hard and deserve some reward. This time *everyone* who substantially improves his score will earn a prize of twenty-five cents, and in addition there will be a prize of $3 for the one who improves the most in the school, $2 for the second best, and $1 for the third. If you work hard to-day, you may win for yourself anywhere from 25 cents to $3. Prizes will be awarded as soon as possible. Do not open the folders until I give the word. Pencils up."

The motive is repeated for Part 2.

The basic score is the achievement the first day. The theory of the test is that the amount of drive induced by the opportunities to work for the Red Cross, for the class, or for oneself will be measured by the accelerated rate of learning on the successive days, allowing for differences in the difficulty of the forms.

In actual practice we found wide differences among the groups first tested with respect to their responsiveness to the different appeals, but the variability among the children of our standard groups, when no motives were used, was so great that the differences associated with the motives were hard to interpret. Many children worked faster and faster without any change in motive, whereas others did less and less on each of the four successive days. The gains from the basic unmotivated score of the first day which we noted in population P could not therefore be taken at their face value, and the test could not be used as a measure of coöperative and unselfish motives.

The Efficiency Coöperation test was later substituted for this phase of our program.

THE SCHOOL KIT TEST

For this a pencil case was used, containing ten articles as follows: drinking cup, pencil sharpener, ruler, eraser, pen, penholder, double pencil, and three other pencils. The kits were of different colors,

red, blue, and green, but only one color was used in any classroom. Each child was provided with a kit, which came as a present from a friend of the school. He was then allowed to give away any part or none or all of the kit in an inconspicuous way in order to help make up some kits for children who had no useful and pretty things of this kind. The pupils' names were written on pieces of paper loosely attached to the large envelopes containing the kits, and the envelopes were numbered in such a way as to identify the children who returned them with their gifts inside. The directions, administered by the teachers, were as follows:

At an appropriate time the teacher will read the following announcement:

"I am requested to announce that a friend of the school has given some school kits to the pupils of certain classes. These are brand new kits (hold up the sample) containing pencils, eraser, ruler, pencil sharpener, drinking cup, etc. These are given to you for your very own, to do exactly as you please with. There is a slip with a name on it on each package."

Pass the kits and give time (10 minutes*) for them to be examined.

"Now I want to pass on a suggestion from the principal. He says there are many schools even in our own country where the children have no pencils or interesting things such as these in these kits. He thought possibly some of you might like to help make up some kits for other children. You might put in one of these little things, or two, or three, or all ten, or the whole thing, box and all, or just the box, keeping the articles for yourself. If you want to help make up some kits for other children, just put whatever of these things you want to give in the red envelope. You may put it all in, or just the empty box, or just one or more of the things, or nothing at all. Tie up the envelope and drop it into the basket whenever you want to before you leave to-day. Please put in the envelope in any case. It is *quite* all right for you to keep any or all of the kit if you would rather. The kit belongs to you."

(The basket is taken to the principal's office at the close of school.)

Note: If a pupil says he wants to give the box to a friend, ask him privately to put a slip of paper in the envelope to this effect.

* The original time was five minutes, but was subsequently increased.

Scoring the Kit Test. As the children gave away some articles more freely than others, we felt that they valued them differently. Hence, we could weight the significance of the giving of any article in terms of the willingness with which the children in general let it go. A large number of penholders, for example, were given away — nearly all the children had fountain pens! But the double pencil, in two colors, was an unusual novelty, and fewer of these were given. We studied the percentages of each article returned and found that we could not make a scale corresponding to the helpfulness scale shown in Figure 7, and could adopt only rough weigntings representing the reciprocals of the percentages of pupils giving away each article.

As second-hand articles were often given away instead of the new ones just received, we gave one-half the score value for these if they were in good condition, allowing only one of each, except in the case of pencils, where a maximum of four substitutes was allowed.

The results of this test, which we shall report in Chapter VI, made us feel confident of its value as a measure of the more strictly charity type of service.

TABLE X

SCORE VALUES OF ARTICLES IN KIT TEST

ARTICLE	SCORE VALUE	
	Population P	Populations X, Y, and Z
Eraser	5	5
Double pencil	5	3
Box	4	5
Cup	4	3
Ruler	3	3
Three pencils (each 3)	9	9
Sharpener	3	3
Pen point	3	3
Penholder	2	2
Maximum score	38	36

In view of the fact that different populations might value these objects differently and also because the boxes were not absolutely uniform, we made a fresh weighting with each new population studied. The differences in the proportion of articles returned were not large, even when there were wide differences in the total amount given away. The score values for population P and those subsequently adopted are shown in Table X.

THE ENVELOPES TEST

Each child is provided with a letter as follows:

Dear Children:

There are sick children in hospitals who need pictures and stories. There are children in some places who never see a bright picture and may have no pictures at all. If you would like to find jokes, interesting pictures, and stories for these children, you now have a chance to do so. Here are four envelopes and in each one you may put something for other children. In each envelope is a card telling what to put in the envelope. If you feel like helping, please follow the directions for each envelope. You may bring the envelopes to the office either all at once or one at a time, as you like.

The first three were long cheap envelopes, about $4 \times 9\frac{1}{2}$ inches. The fourth was a large heavy envelope, about 9×12 inches. The slips were as follows:

Envelope 1: Put in this envelope one or more jokes to make joke books for hospital children. Bring your envelope to the school office. If you would like to help put such books together when the jokes are in, sign here and put this slip in the envelope.

Sign only one of the statements.

1. I will help make some joke books when the jokes are collected.

Name_____

2. I would like to help make the joke books but I shall not be able to do so.

Name_____

Envelope 2: Find some interesting puzzle pictures which you think that children in the hospital might like to have. Put these puzzle pictures in this envelope and bring it to the school

office. If you would like to help put these pictures in little books for the sick children, write your name here and put this slip in the envelope.

Sign only one of the statements.

1. I will help make some puzzle picture books when the puzzles are collected.

Name_____

2. I would like to help make some puzzle picture books but I shall not be able to do so.

Name_____

Envelope 3: Put in this envelope a short story to be read to a sick child. Either write the story yourself or copy it out of a book or magazine with any changes you may care to make. It should be interesting and short. Tell where you found it. Bring the envelope to the school office. If you would like to help collect these stories into little books, sign your name here and put this slip in the envelope.

Sign only one of the statements.

1. I will help put the stories into books when the stories are collected.

Name_____

2. I would like to help put the stories into books but I shall not be able to do so.

Name_____

Envelope 4: Put in this envelope a beautiful picture. It should be a copy of some famous painting. It would be of help to them if it were pasted on a piece of cardboard so that it will not get broken. The picture will be sent to children in mountain schools to decorate their schoolrooms or homes. Bring the envelopes to the school office.

The teachers were also instructed as follows:

After the envelopes are passed out nothing should be said about the matter for three days. On the third day the teacher will say: "A good many envelopes with puzzles, pictures, etc. have been left in the office. If any of you are planning to bring your envelopes in, there is still time to do so."

On the seventh day the teacher will say: "I am asked to remind you of the envelopes you took home in case any of you have forgotten about it. If there are any who still want to bring something, there will be a few more days."

On the tenth day the teacher will say, "The last day for bringing in material in the envelopes — puzzles, pictures, etc. — is next _____ (naming four days later)."

Nothing more should be said about it.

Extra sets of envelopes are available in the office for any who may lose theirs.

Scoring the Envelopes Test. In scoring this test, it was assumed, as for the kits, that the value of an article turned in was in inverse proportion to its prevalence. But here also, we realized the need of adjusting our scores to the populations dealt with in view of the fact that families in some communities would be in a better position to supply the clippings and pictures than would families in some other communities. Also there were common-sense considerations, such as the availability of material as compared with the trouble taken to prepare it. Hence double value was given for mounting a picture or for writing out a story, joke, or puzzle instead of merely cutting it out.

On this basis our score values for school P were as follows:

	Clipped	Written or Mounted
1. Jokes	3	6
2. Puzzles	5	10
3. Stories	10	20
4. Pictures	5	10

This gave us an actual range of scores of 0 to 150.

Our results with the Envelopes test were unsatisfactory in the sense that a large proportion of children did nothing at all. But we felt that some attempt to have the pupils work at a project of this kind at home instead of at school was necessary and so we retained it. Our later results made some modifications in the score values of the various articles necessary. By comparing the proportion of objects of each sort turned in, the proportion of the population handing in at least one of each kind, and the average number of each kind handed in, we arrived at the set of score values shown in Table XI.

These score values are of course only estimates. We graded them in such a way that a large number of one object of low scale value would not pile up a score greater than that of the more highly scaled objects. The

table shows, therefore, the value of *one* of each object, and also the value of each successive number of similar objects up to a total of five, after which no additional values are assigned, no matter how many objects are turned in. This was the method used in all subsequent testing.

TABLE XI

SCORE VALUES OF ENVELOPES TEST FOR POPULATIONS X, Y, AND Z

OBJECTS	NUMBER OF OBJECTS BROUGHT				
	1	2	3	4	5
Jokes clipped (cut out)	5	9	12	14	15
Jokes copied	10	17	21	23	25
Puzzles clipped	5	9	12	14	15
Puzzles copied (or cut up)	15	22	26	28	30
Stories clipped	5	9	12	14	15
Stories copied	15	27	36	42	45
Pictures unmounted	5	9	12	14	15
Pictures mounted	10	18	24	28	30
Maximum score					190

It will be recalled that the children had the opportunity of saying whether they would, or would if they could, come to the school and help make books of pictures, puzzles, etc., for sick children. No use could be made of this score in the measurement of individuals as there were too many conflicting engagements for the time set to come. We invariably had some assistance from interested pupils, however, as might be expected, and in some cases the differences among the classes with respect to the percentage saying they would come if they could was of some value to us.

ADDITIONAL SERVICE TESTS

In addition to the tests tried out in population P two others were devised for use with populations X, Y, and Z. They were administered as a single technique,* combining and developing the possibilities inherent in the Self-or-Class test and the Learning Exercises.

* For the development of this procedure Dr. J. B. Maller is responsible. The directions for administration are printed in Appendix B.

As the interest of the test lay in the variation induced in behavior as the result of contrasting motives — selfish and unselfish — it was necessary to elaborate and standardize the activity which was to be affected. For this purpose simple additions were used, as we had already found this activity to be reasonably constant when no change in motive was introduced.* Booklets were prepared containing twenty sheets, on both sides of which were printed, in varied order, the examples shown on page 69. The sheets were numbered in such a way as to enable us to identify each sheet of each child once his name was placed on any one of them.

The Efficiency Coöperation Test. In its final form the test was administered in such a way as to alternate work for self in a contest with other individuals with work for one's class in an interclass contest. One minute was spent on each side of each sheet, giving thus two minutes to work for self followed by two minutes of work for the class; and this again alternated with two for self, two for class, etc. Twelve sheets were used in this way, six units for each motive.

The Self Motive. The self motive used was one of ordinary competition for recognition and reward. The subjects were told that they were going to have a speed contest "to find out your speed of work in additions. This test will tell us who of you is the fastest worker, the second fastest, third fastest, and so on to the very slowest worker in this class. Prizes will be given to those who do fast work. Every one of you has a good chance to win a prize. When I say, 'Go,' you will start to add and continue until I tell you to stop. The faster you work, the higher will be your score." The pupils wrote their names on each of these sheets before starting to work.

The Group Motive. The group motive used consisted in creating a contest between classes. It was similar to the intraclass competition used to measure the self motive. "This is to find out which class in this school is the fastest, second fastest, third, fourth, and so on to the very slowest class. Class prizes will be given to those classes which do fast work. This class has a good chance

* See *Studies in Deceit*, Book Two, pp. 32 ff.; also Picture Inhibition test.

Score.....

Do each example as you come to it, moving across the page from left to right.

6	9	7	5	7	2	3	4	2	5
8	2	9	8	6	6	5	3	8	9
—	—	—	—	—	—	—	—	—	—
4	6	4	9	3	3	7	2	6	6
2	8	9	8	6	7	5	3	9	5
—	—	—	—	—	—	—	—	—	—
3	8	8	5	2	8	7	6	7	6
9	4	7	7	5	3	4	5	3	9
—	—	—	—	—	—	—	—	—	—
4	8	7	7	8	3	2	5	9	6
3	5	2	8	6	8	4	3	7	4
—	—	—	—	—	—	—	—	—	—
6	8	3	5	2	7	8	5	9	7
3	9	2	4	6	9	4	2	3	5
—	—	—	—	—	—	—	—	—	—
2	6	8	9	4	9	2	4	6	5
7	7	2	2	6	4	9	8	2	8
—	—	—	—	—	—	—	—	—	—
6	9	3	2	4	5	2	9	4	8
7	5	5	8	3	9	4	2	9	9
—	—	—	—	—	—	—	—	—	—
7	3	5	8	6	5	4	4	7	8
3	6	7	5	9	6	3	5	8	6
—	—	—	—	—	—	—	—	—	—
2	3	6	3	5	4	6	9	8	7
5	8	4	9	3	2	3	6	7	2
—	—	—	—	—	—	—	—	—	—
2	9	4	7	6	7	4	7	9	5
3	7	7	3	5	4	8	2	4	7
—	—	—	—	—	—	—	—	—	—

FIGURE 10

FACSIMILE OF ONE SIDE OF SHEET USED FOR THE COÖPERATION TESTS

to win a prize. The classes that will not try hard will naturally be at the bottom of the list of the classes in this school." The children were told not to write their names on these sheets but to write the name of their respective class, as 7A, etc. " The score of your class will be the number of examples all of you do for your class."

The child's score for self was the average number of examples done in the twelve minutes devoted to working for self, and his " class " score was the average number of examples done in the twelve minutes devoted to working for the class. The difference between these two averages was the Efficiency Coöperation score. If the subject worked as hard for his class as he did for himself, his coöperation score was zero. If he did more for the class than for self, his score was positive. If he did better work for self than for class, his score was negative.

In order to give all scores the same sign and have increasing scores represent increasing amounts of coöperation or class loyalty, each child's raw score was added algebraically to 100. If he did more for class than for self on the average, his score was over 100. If he did more for self than for class on the average, his score was less than 100. If he worked under the two motives with equal efficiency, his score was just 100.*

The Free Choice Test. After the Efficiency Coöperation test was finished, the examiner said: " There are seven more sheets. You may work on as many of these as you wish for yourself and on as many as you wish for your class. You may even do all the tests for yourself or all of them for your class. Whatever you do for yourself will add to your own score and whatever you do for your class will add to the class score. At the end of the test I will collect all the papers at one time. Look at the next sheet. Remember, if you write your name on this sheet, it will count for yourself and for your own score. If you write the name of your class on it, the additions will count for your class. When I say, ' Go,' write either

* In this volume we use 10 instead of 100 as the basal number in the case of mean scores, as the range of means was rarely more than 90 to 100, whereas the range of individual scores was 70 to 120.

your name or the name of your class on the sheet and begin to add."

Only one side of these sheets was used, as the primary purpose was to find the proportion of units given by the pupils to the class and the proportion kept for self. The directions were repeated for each sheet. As there were seven sheets in all, the scores ranged from 0 to + 7 in accordance with the number kept for self and the number given away. The odd number compelled the expression of a preference for self or class, as the total could not be divided equally between the two.

The distinction between this test and the Efficiency Coöperation technique lies in the mode of expressing one's interest in self as contrasted with interest in one's group. The first test involved no self-sacrifice. There was no inherent conflict between the two motives in the test itself. Equal opportunity was given to work *both* for self and for class. The group motive functioned in precisely the same way as the self motive — to get work done. As fatigue or ennui set in, the amount of work tended to decline. If it declined faster for the units devoted to class than for the units devoted to self, the score came out in favor of self.

On the other hand, the Free Choice test set the two motives in open conflict. It was a case of "either, or." Each child chose whether he would give work to the class or keep it for himself. Self-sacrifice was involved, for whatever he gave away he could not have for himself.

In addition to their value for our scheme for testing coöperation, in which one motive was set in conflict with another, these tests also represented very well our general theory of conduct testing. As will be recalled, in our *Studies in Deceit* we used an analogy from physics to help explain the specific nature of conduct tendencies. We found that a single type of deceptive conduct was accompanied by or represented specific amounts of attitude or drive. This attitude we attempted to measure by the amount of work it would do or the amount of resistance it would overcome. So in the Efficiency Coöperation tests, we set up a uniform resistance in the form of a monotonous and fatiguing activity and measured

the amount of work done under the two contrasting motives — self and class.*

THE TOTAL SERVICE SCORES

In order to determine the most satisfactory way to combine all the service scores in a single total, we correlated each with all the others, noting also their variability and reliability.† Their inter-*r*'s, however, were all too low to be of any assistance in this computation and we depended on SD's and observation in making our combinations.

We made first two subtotals. One of these consists of the Efficiency Coöperation, Free Choice, and Money Vote scores, and the other, the Kits and Envelopes. In the first we gave less weight to the Money Vote test, as we had no measure of its reliability and it represented a vote rather than an independent individual act.‡ In the second subtotal we gave equal weight to the Kits and Envelopes.§

The grand total, always referred to as the total service score, was found by adding these two subtotals, giving the same weight to each.‖

SUMMARY

Our service tests, as used in our major populations, X, Y, and Z, employ five separate techniques, each of which consists of one test containing several minor elements. In three of these tests the factor of class loyalty enters, namely, in the Free Choice, Efficiency Coöperation, and Money Vote. In the first two of

---

* The details regarding the development and standardization of these tests are discussed by Dr. Maller in his monograph "Coöperation and Competition," *Teachers College Contributions to Education*, No. 384, 1929.

† The problem of reliability is discussed in the next chapter. The intercorrelations are given in Table XXIII.

‡ The sigmas of these tests are: Efficiency Coöperation, 5.8; Free Choice, 1.9; Money Vote, 2.4. In adding, the raw scores were used, with the Free Choice score doubled. The sigma of the combination is 8.5.

§ The sigmas of those coöperating (omitting zeros) are: Kits, 4.6; Envelopes, 2.0. The sum of the scores for Kits plus twice those for the Envelopes gave a sigma of 8.9.

‖ As the Kits-Envelopes sigma was 8.9 and the sigma of the other subtotal 8.5, we simply added the former combination to the latter combination. The total scores range from 60 to 160, with a sigma of 12.3.

these, the highest scores go to those who are most devoted to class or group interests as over against self interests. In the third, the class interest is only one consideration, as in voting what to do with their prize money the children have also to consider giving it to something for the school as a whole or to some philanthropic cause outside the school. The last of these choices receives the highest credit, the school cause second, and the class cause third, while self interests, of course, come last.

The two remaining tests have no element of group loyalty in them but relate directly to charitable enterprises in which the children engage as individuals. These are the Kits and Envelopes tests. The first is a classroom test and the second involves the homes, as the material to be collected for hospital children is not provided by the school.

Reference to our theoretical discussion of coöperation and charity in Chapter II will reveal how far we have covered the ground which is there described. Obviously only a beginning has been made of a study of the tendency to serve.

CHAPTER IV

REPUTATION FOR SERVICE

One of the most fundamental and at the same time most difficult of all the problems of character testing is the problem of validity. What assurance have we that our so-called service tests measure the type of behavior tendencies classified under this title? How well do the tests we now have represent the entire field of altruistic conduct? As in the case of our tests of deception, we need to remember that we are concerned for the moment with the outward aspects of conduct rather than with its motives. We place our subjects in life situations and keep an objective record of what they do. Each act, therefore, is just what it appears to be. The subject does not help at all or, if he helps, he does something which can be counted or otherwise measured in terms of the amount done. The consistency of his behavior in such matters is shown by the proportion of opportunities to help which he utilizes or by the total amount of work done. In a sense, then, the tests require no validation. They measure what they seem to measure, by whatever name it may be called.

Another way of thinking of validity, however, needs to be taken into consideration. In the early days of intelligence test construction, it was customary to attempt to validate tests by comparing their results with some external criterion. Binet depended on teachers' ratings in developing his famous series of individual tests of intelligence. The psychologists who made the Army Intelligence tests depended largely on officers' ratings for evidence of the validity of the tests. Experience has shown, however, that the ratings usually used for validating tests are themselves unreliable sources of information. The tests are felt to be superior to the

criteria used to validate them, and so the present tendency is to depend for validation on evidences of internal consistency and a careful analysis of how the test hangs together.

When ratings and records can be shown to be reliable, however, they continue to be of considerable importance. They constitute a measure of *reputation*, and reputation is significant for any discussion of character. To be sure, the child's inner character does not necessarily correspond to measures of his reputation any more than it necessarily corresponds to measures of his external conduct, but it is important for us to know the *extent* of agreement between reputation and conduct if for no other reason than the light this fact may shed on character itself.

In spite of the inadequacy of any easily obtainable external criterion to serve as a measure of the true significance of a character test, for the reasons just stated we have felt it wise to gather all the information we could concerning the reputations of our cases. We have devised several ways of discovering what it is for each child and of determining its amount and value, namely, (1) the record of social service in school projects, (2) the Portrait Matching device, (3) the "Guess Who" test, (4) the Check List, and (5) the Conduct Record. The first and third were used in population P only, and the second in P and certain groups in X, Y, and Z. These will be described in order.

THE RECORD OF SOCIAL SERVICE

Our experience with records of this sort, which we reported in Chapter I, convinced us of the value of this procedure. We changed the plan somewhat, however, so as to give the teachers the opportunity of recording the pupil's behavior as it occurred. The following memorandum was given to each of the teachers of population P who had signified her willingness to coöperate, after the matter had been carefully discussed.*

* Records were returned by the fifth- and sixth-grade teachers, accompanied in one or two instances by very full descriptions of class projects.

Memorandum to Teachers concerning Recording of Helpful Acts

This memorandum may be of some assistance for keeping in mind the things that are needed in the way of a record:

1. Record only helpful acts done. Do not attempt to record attitudes or motives. They will be taken care of in another way later.
2. Record only such acts as every child has an equal chance to take part in.
3. Make all records quantitative, that is, count the times that acts are performed, or quantity of achievement, or length of time spent. If this is not possible, then indicate on a scale of zero to four the *relative* amount of work done by each pupil, letting two be the average amount for the deed.
4. Keep the record going until the end of the spring session.
5. Keep the record only in case the facts are known, but make some record for each child for each type of activity. When nothing is known, cross out the square concerned entirely by drawing a large X from corner to corner.
6. Let each vertical column stand for some particular activity and where possible let these activities be grouped within some major project which has some clear-cut identity in the minds of the pupils, such as the acts centering around the Christmas festival.
7. Include as much as possible from memory for the period from the beginning of the school session to date.
8. Date the time occupied by each activity when labeling columns concerned.

If you have any questions, we should be glad to talk over the plan at any time.

These records were begun at Christmas time and were continued to the end of the school year. Each teacher was supplied with a list of her pupils and spaces for naming each activity and stating the amount done by each pupil. The records were regarded as representing actual facts and were summarized so as to give the pupils comparable service scores in terms of their opportunities.

The following section from one teacher's record will illustrate how the final score was reached:

Names	1	2 Nursery			3	4	Score (Average of 1+4)
	Christmas Giving	Coin Cards	Motion Pictures	Candy Sale	Framing Pictures	Average of Nursery	
J. G.	2	2	4	4	0	3	25
M. O.	4	1	0	2	2	1	25
T. L.	2	0	1	1	2	1	15
F. A.	2	2	1	2	4	2	20

In computing the score, a rough average for each project was found, and these averages were again averaged in order that each project might have the same weight in the total. In this case the picture framing episode was a purely class project and so was not included in the service score. J. G. has ratings of 2, 4, and 4 on the nursery project, which roughly average 3. He has a rating of 2 on the Christmas project. These two then average 2.5 or 25, which is taken as his service score.

As these were records of fact, no problem of either reliability or validity was raised as the only way to increase the significance of the record was to gather more facts, and we already had all the facts available on this type of behavior.* The scores were well distributed, with only four pupils getting the highest possible score and five the lowest possible, the average falling at just two; that is, 167 pupils took advantage, on the average, of two chances in four to be of help.

The service record was used in population P partly because of the unusual amount of social work done by the pupils of this school and partly because the teachers were well informed about most of the children. Under such circumstances it made a useful criterion, but the conditions in X, Y, and Z did not warrant our using it with these groups.

THE PORTRAIT MATCHING DEVICE

This scheme was first suggested by the vivid character sketches of Theophrastus. Picking out some quality, such as meanness, he

* Intercorrelations and predicted validities for this type of record are reported in Chapter I. We shall take .80 as the present reliability coefficient

gathered into a single picture all possible ways in which this sort of behavior is commonly manifested and attributed these habits to one person. The result is, of course, a caricature, but no one could mistake the intention; and, as so often happens, the overdrawn picture seems to fit persons of one's acquaintance extraordinarily well.

We attempted, therefore, to build up sketches of boys and girls who exhibited the tendency to be of help to others in various degrees from a complete absence of any tendency whatsoever to the largest possible amount of helpfulness that one could imagine.

In order to get material for the sketches, we first analyzed helpful behavior into a series of objects, motives, and resistances. The following three lists resulted:

Objects

A member of the immediate family in need
A friend of his
A fellow pupil in need
His teacher
A class enterprise
A school enterprise

Some younger child in trouble
An animal that needed help
Children's hospitals (local)
Day nurseries or help for babies (local)
Orphan homes (local)
Day nurseries, or food for babies, or help for babies in other lands
Anyone who needed help
Orphan homes or children's hospitals in other lands — children same ages as helpers
Anyone who asked for help
A stranger
Schools in other lands — students older than the helper
Mission schools in America — students older than the helper
Homes for the aged (local)
Hospitals for adults (local) or such institutions as Mills Hotels for the unemployed

Homes for the aged — in other lands
Hospitals for adults — in other lands
A foreigner here
One who did not deserve any help

Resistances

To give a toy or article of which the child is tired or for which he has no further use

To leave an activity of a given interest and turn to another of equal interest

Even if no one knew about it

Even if the person he helped did not thank him

To give up time that would otherwise be devoted to school lessons and make up the school work later

To give a small percentage of one's allowance or earned money which would otherwise be put in the bank

Even if it took a lot of time

To earn money to give away

To engage in an activity of moderate interest which required some extra effort, such as getting to school an hour early

Even if he got tired helping

To leave a very interesting and engaging activity to take up a distasteful task and continue at it until thoroughly fatigued

Even if he had to do something wrong in order to do it

Even if parents or teachers forbade it

To give up some cherished object which has been obtained with considerable effort and is not easily replaced

To give up money that has been saved to buy some object or some kind of pleasure which the child has long wanted and is now about to realize

Even if he might be seriously hurt himself

Even if laughed at or scorned by fellow pupils

Even if imprisonment is risked

Motives

If he got paid for it

If there was a prize for the one who helped most

To receive great applause from fellow pupils and adults — become a hero

To get something in return, either objects, help, or money from the one helped
To receive the approval of fellow pupils
If there was a contest to see who could do the most
If his best friends were in on it
If the rest were helping too
If the activity involved is more fun or interesting than alternative activities in which he would otherwise engage
To receive approval of parents and other adults
If the teacher asked him to
If the principal asked him to
If his folks wanted him to
To leave one activity and turn to a slightly more interesting one

---

If his sympathy, or some other emotion, has been aroused temporarily
If he wants to be helpful
If he has a genuine interest in promoting the happiness and welfare of all people
If his conscience would bother him unless he did

Each item was printed on a card, and we then had seven judges arrange each list in an order representing degrees of appeal. The lists are printed here in the order which resulted from these judgments. The objects thought to be most likely to elicit helpful behavior come first. The first six are admixed with loyalties of various kinds and were not included in the portraits. The presumption is that those who would help the objects toward the bottom of the list have a greater tendency to be of help than those who would confine their assistance to objects higher up on the list, which seem to be inherently more appealing.

The resistances are printed in order of size. Those who are willing to overcome the resistances that come toward the bottom of the list are presumably more helpful than those, for example, who would not be willing to earn money or who would refrain from helping if the beneficiary would not thank him for his assistance.

All but the last four motives are extraneous and begin with the strongest — that is, those who would help only if paid for it are thought to be less generous than those who would help even if there

was no more reason for doing so than that the proposed activity is slightly more interesting than what he would otherwise be doing. The last four motives are stated as directly related to helpfulness, in the form of a natural sympathy or an ideal of helpfulness or a desire to make everyone happy or a sense of duty.

Having now a rough scale of the constituents present in every-day acts of helpfulness, we made up typical situations, combining the factors listed in such a way as to represent characters of increasing altruism.

These paragraph portraits were then submitted to forty-eight judges, who were asked to arrange them in order from the least helpful to the most helpful. There was considerable overlapping at certain levels, as may be seen from Table XII, which shows the way in which the forty-eight judges arranged the pictures. Cases D, H, and V, for example, are apparently much alike.

TABLE XII

Evaluation of Original Ten Portraits*

Portraits	10	9	8	7	6	5	4	3	2	1	Mean Position
T	45	3									9.9
F	3	37	6	2							8.9
N		7	19	17	2	3					7.4
R		1	22	19	2	4					7.3
D				7	12	12	12	2	3		5.0
H				1	19	12	7	9			4.9
V			1	1	12	11	14	6	1	2	4.6
L					1	3	8	19	10	7	2.9
J				1		2	5	9	22	9	2.44
P						1	2	3	11	31	1.56

This table reads: 45 judges put portrait T at the top of the list and 3 judges put it in second place; 3 judges put portrait F at the top, 37 rated it next to the top, 6 gave it third place, 2, fourth place; etc.

* A rough approximation to a scale was found as follows: Case P was taken as representing zero helpfulness. Nine judges, or 19%, placed J also at this level or failed to distinguish between J and P. Case L has a mean score of 2.9, which is third from the bottom. Seventeen judges, or 35%, placed it sec-

Case H was consequently omitted and Case B substituted to give a greater variety of portraits at the upper levels. The final set, in order of helpfulness and with the score values based on the distribution of the pictures by the judges, is presented here:

9 — T. T is sincerely interested in promoting the happiness and welfare of everyone. His warm spirit of friendliness extends to all, no matter where they are or what their race, class, or creed may be. Even the slightest need stirs him to some friendly act. So strong is this purpose that he would endure the scorn of his fellows in order to help a struggling cause or to help persons in trouble, even though they might not seem to deserve help of any kind.

7 — F. F is quite sincere in wanting to be of help. To him it is not so much a matter of high purpose as of plain decency. When the cause seems to be a worthy one, he is very generous and would give a cherished possession or deny himself long anticipated pleasures even for such seemingly remote interests as scholarship funds for older students in other lands or for hos-

---

ond or first. The sigma values of ½ the successive per cents, together with derived rough scale values, are given in the following table. The actual values used are a compromise between the sigma scale and the successive means shown in Table XII.

Sigma Values of Percentage Overlap		New Order and Scale Values of Portraits		
Portraits	Sigmas	Portraits	Scale Values	Actual Scores
T	1.89	T	9.5	9
F	1.41	F	7.6	7
R	1.56	B (new)		7
N	1.65	N	6.2	5
D	.93	R		
V	1.34	D	4.6	2
H	.96	H		
L	.93	V	3.6	2
J	1.37	L	2.3	1
P	0.00	J	1.4	1
		P	0.0	0

pitals for adults. He would persist in his helpful acts even though people thought he was very foolish to do so. He is prompt to offer his services to anyone obviously in trouble.

7 — B. B is always thoughtful for others around him and sees many opportunities for little kindly acts without waiting to be asked. On occasion he would take great risks to help anyone in danger, and he always makes his own convenience and pride a secondary consideration. He is not greatly interested, however, in remote needs unless they are serious and appealing.

5 — R. R has been taught that it is his duty to help, and so his conscience bothers him if he doesn't do what he regards as his proper share. He feels that worthy causes should be helped generously and his sense of duty leads him to do tiresome and disagreeable work or take trouble to earn money for objects generally regarded as needing public support. There is doubtless a limit to what he would give, such as a tenth of his income, but he has no thought of himself or his convenience in doing his full share as he sees it, and of course very appealing objects arouse his sympathy.

5 — N. N does not think much about being helpful as a general thing, but he has a kindly nature and his sympathies are easily aroused. When his emotions are appealed to, he is ready to help almost anyone or any cause, but would probably draw the line at helping people he disliked or who seemed foreign or undeserving. He always lends a hand when the occasion is obvious and, if the appeal is strong enough, would give away his own things or money he had saved for himself in order to help even such objects as a college for an underprivileged section of the country or a hospital for poor people. The disapproval of his parents or teachers would not dissuade him.

2 — D. D is not hard-boiled by any means, but it does not occur to him to go much out of his way to help anyone nor could he be persuaded to do much of anything for objects not naturally appealing. He likes to help along when children or animals are involved or in case of very great need, as in famines or disasters, for which he is willing to earn extra money if the work is not too hard. He gives up some of his things quite readily for such causes, or takes time from his play. And for less appealing causes, where another's sense of duty might be invoked, he

would be available for help if he thought that he might get a "stand in."

2 — V. V is entirely conventional in what he does to be of help. The obvious things, which appeal to the vast majority, he chips in on, in a more or less perfunctory fashion. His sympathies are rarely stirred, but he realizes that a certain amount of philanthropy is expected of him and wishes to be counted in as contributing to good causes, even at the expense of play time or part of his allowance, though he does not exert himself particularly.

1 — L. L is not likely to help anyone at all unless he finds that others are doing so, especially his best friends. He'd chip in a little for flowers for a classmate or help his elder brother take care of his small sister if it did not take too much time from his play or if he didn't have to do too much work or spend more than a few cents. But for the most part he doesn't want to help, nor does he care whether he is regarded by others as helpful.

1 — J. In only extreme cases of need is J found doing anything for anybody, and even then he does so only if the rest try to get him to do something, in which case he may, for example, bring some discarded toy or garment for a child in distress or help others make him a toy if he enjoys making things of that sort very much. He would not care about being thanked or praised, but he would hope for some reward, such as a dime from his father or a good mark in Heaven or a star in his crown or an access of grace.

0 — P. Here is your thoroughly hard-boiled youngster. The person or cause needing help makes no difference at all to P; but if someone would call for it, he would let him take away some object he wanted to get rid of provided he would get some thanks or reward or prestige. Or he would leave something he didn't want to do anyway in order to engage in some helpful activity if it were very interesting on other grounds and if people would applaud him for doing so. But he does not care two cents about the object or about helpfulness as a duty or as the proper thing or as a way of winning favor or grace.

The portraits were used as a kind of rating device the steps of which are elaborately defined and of known size. The teachers

did not know the order of steps, however, and they were encouraged to use the pictures as one would use a handwriting scale, taking each child and matching him against a portrait. Of course, there remained a considerable subjective factor, but we felt that the tendency to rate a child high in helpfulness merely because he had a good reputation on other counts was in part overcome. The directions used by the teachers were as follows:

The ten paragraphs are little pictures of people. You probably know persons just like them among your acquaintances. You could match each picture with at least one person. Some of the pictures would fit a great many people. The pronoun "he" is used to refer to either sex.

Procedure: The matching of portraits and children need not be done at once. It would be well to read the set of portraits to yourself on several days before attempting any matching. When you come to the actual work of matching, you may find that many children in your class will not fit any portrait, or that you do not know enough about many children to attempt to match them with portraits. Some portraits may fit several children. Some children may fit several portraits. All this is to be expected.

Matching: Read the first portrait.* When you have this picture clearly in mind, turn to the page of names and check in column D as many names as the first portrait fits. If no one fits D, put down no checks.

Then read the second portrait. With this clearly in mind, turn again to the names and check in column R as many as R fits. If R fits any pupil already checked in column D, do not hesitate to make a check also in column R. If R fits no one in your class, put down no checks. Any name may have from no checks at all to ten checks if all ten portraits fit.

Do the same with all the other portraits.

Then lay the completed portraits to one side.

Later on, check your first records as follows:

Take each *name* and, with this pupil in mind, go over the portraits and see which one or ones, if any, fit this pupil, placing your checks in the columns of the carbon copy of names without reference to the first list previously checked. When you have finished doing this with

* As presented to the teachers, the portraits appeared in the following order: D, R, L, B, V, P, N, T, J, F.

each pupil, compare the results. If differences occur, you may consult with others about discrepancies and get their judgments on the particular children involved. But remember that children relatively colorless or unknown may quite properly fit no portrait at all.

Following these instructions, the teachers frequently matched one name with more than one portrait. In these cases, the portrait score of the child was the average of the values of the portraits used.

Value of the CEI Portraits. We attempted to determine the value of the matching device in several ways. First we noted the tendency to use only the finer or more coöperative portraits. This is indicated by the distribution of each rater's scores, which is given in Table XIII.

TABLE XIII

DISTRIBUTION OF SCORES ON PORTRAITS — POPULATION P

SCORES	TEACHERS						
	1	2	3	4	5	6	ALL
9	1	0	0	0	0	0	1
8	1	0	1	1	0	0	3
7	4	3	8	5	7	5	32
6	0	7	4	1	2	2	16
5	4	1	10	1	3	1	20
4	1	16	0	1	0	0	18
3	0	0	2	1	0	0	3
2	6	2	4	9	9	6	36
1	5	0	1	3	6	4	19
0	0	1	0	0	0	0	1
Number	22	30	30	22	27	18	149
Mean	3.91	4.53	4.10	3.73	3.70	3.78	4.19
SD	2.61	1.54	1.87	2.42	2.44	2.51	2.28

This table reads: Teacher 1 matched one child against portrait T, thus giving him a score of 9; three teachers each assigned one child to portraits which averaged 8; teacher 1 gave 4 pupils scores which averaged 7; etc.

The scores are obviously well distributed. The children are not all put high up on the scale, as so often happens in ratings. There is even one case matched against portrait P, scoring zero.

We secured a measure of reliability by having each teacher use the scale twice, as indicated in the directions. We found that the two sets of scores agreed reasonably well, which may be taken as evidence that they do not express merely a passing whim of the rater.*

The portraits required a great deal of time, as well as intimacy with the pupils. Consequently we were obliged to omit the use of this technique in several classrooms, while the records in others were incomplete and therefore were not utilized. For population P, however, this served as one of the best of our measures of reputation.

THE "GUESS WHO" TEST

This device is really an abbreviation of the Portraits device, except that in our actual use of the instrument we included several other types of behavior besides charity and arranged the material so that the pupils could do the matching instead of the teacher. We shall print here, however, only those portions of the scale which have reference to the tendencies we are now studying.

The "Guess Who" Test †

Here are some little word-pictures of children you may know. Read each statement carefully and see if you can guess whom it is about. It

* The *r*'s are given in Table XIV.

TABLE XIV

RELIABILITY OF PORTRAIT MATCHING — POPULATION P

TEACHER	SELF *r*	PE
1	.967	± .015
2		
3	.838	± .040
4	.954	± .020
5	.507	± .100
6	.934	± .025
Average	.840	± .020

† The "Guess Who" test is copyrighted by the Institute of Educational Research, Teachers College, and is quoted by permission.

might be about yourself. There may be more than one picture for the same person. Several boys and girls may fit one picture. Read each statement. Think over your classmates and write after each statement the names of any boys or girls who may fit it. If the picture does not seem to fit anyone in your class, put down no names but go on to the next statement. Work carefully and use your judgment.

1. Here is the class athlete. He (or she) can play baseball, basketball, tennis, can swim as well as any, and is a good sport.

_____ _____

_____ _____

2. Here is the best all-around student in the class. He (or she) always gets his lessons and makes a good recitation. He knows what he is talking about.

3. Here is someone who is always ready to play or work with the rest even when he (or she) can't have his own way.

4. This is a jolly good fellow — friends with everyone, no matter who they are.

6. This one is always picking on others and annoying them.

9. Here is someone who always wants his own way. This person wants to be boss and have everyone do as he says.

10. Here is someone who is kind to younger children, helps them on with their wraps, helps them across the street, etc.

12. Here is someone who will always stop what he is doing to help a fellow pupil. This person is never too busy to lend a hand when someone is in trouble.

15. This one would never go out of his way to help anyone or take any trouble to be of use to others.

18. Here is a crabber and knocker. Nothing is right. Always kicking and complaining.

19. Here is someone who is always doing little things to make others happy.

26. Here is a very selfish person, who will not share anything. He always wants to know what he can get for himself and never thinks of others.

The first two items were intended as buffers to establish the right attitude toward the remaining questions. The sheet was called a test to stimulate careful thinking regarding the facts. The

persons filling out the blanks did not put their own names on them, as this might have put some restraint on the pupils and made them hesitate to make use of the portraits of less desirable characters.* In one instance, however, the writers' names were inadvertently secured, which gave us much valuable information regarding the behavior of individual pupils under these conditions. These and other interesting features of this technique will be taken up in Volume III, where we shall discuss the problem of a general character criterion. Our interest for the moment is confined to evidences of coöperative tendencies, and this device seems to provide a way of getting at the reputation a child has among his classmates.

* In a later experiment, 117 children in three classes were asked to write the names of the most coöperative, the next most coöperative, and the third most coöperative child in their class, and then the names of the least coöperative, the next, and the next. The first time this was done they did not write their own names on their ballots, but they could be identified in another way. The next day they were asked to do this again, their own names being already on the ballots. The difference between the way they voted on the two days is shown in Table XV

TABLE XV

Effect of Signing One's Name on a Ballot

	When the Rankers Put Their Names on the Ballots	When No Names Were Given
Mean number of positive votes cast by each child	3.74	1.47
Mean negative votes	1.18	1.04
Ratio of positive to negative votes	3.17	1.41
Positive minus negative mean votes	2.56	0.43
Percentage of children casting a positive vote for themselves	32%	74%
Percentage of children of whole group receiving positive votes except their own	41%	67%

This table seems to indicate that, when individuals rank one another and are told to put their name on the ballot, they tend to cast more positive votes, to center the positive votes on a few, and to vote less often for themselves.

As the teachers also filled out the blanks, we were furnished with a check on the pupils' ratings of one another.

Value of the "Guess Who" Test. Since any child might have been voted for on each of the five positive or five negative items which we have just now quoted as bearing on altruism, or marked on all ten if he were of such a chameleon complexion as to fit all the pictures, or by half the pupils on the positive items and by half on the negative if he made a different impression on different children, his total score might range from a maximum negative score of five times the number in the class, through zero, or an even balance of plus and minus scores, to a maximum plus score of five times the number in the class. In practice, we found distributions such as those shown in Table XVI.

TABLE XVI

DISTRIBUTIONS OF "GUESS WHO" SCORES — POPULATION P

SCORES	GRADE 5			GRADE 6			TOTAL
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	
+ 26	1						1
+ 22		3					3
+ 18		1			1	1	3
+ 14		2	2	2		1	7
+ 10	3	1	1	1	1	1	8
+ 6	3	9	3	4	7	4	30
+ 2	4	6	13	7	9	9	48
− 2	6	4	7	9	6	7	39
− 6	6		3	2	1	3	15
− 10	1	1			2	2	6
− 14		1		1			2
− 18+		2				1	3
Number	24	30	29	26	27	29	165
Mean	1.2	3.6	1.8	1.2	1.6	0.8	1.8
SD	7.6	10.8	4.8	5.8	5.6	6.4	6.1

The scores shown in Table XVI are the number of positive votes, that is, the number of times a name appears in items 3, 4, 10, 12,

and 19, minus the number of negative votes, or the number of times a name appears in items 6, 9, 15, 18, and 26.*

The total distribution shown in Table XVI is quite symmetrical and indicates the absence of the tendency to give different pupils the same score, which is a common defect of rating schemes. This willingness to use the low end of the scale, or the " Guess Who " items representing undesirable behavior, is still more clearly shown by the fact that the average number of positive votes received by the pupils was 5.6 and the average number of negative votes received was 4.6.

To secure a measure of the reliability of the " Guess Who " device, we divided the five positive and five negative scores into two sets, one containing items 4, 6, 19, and 26 and the other, items 3, 9, 10, 12, 15, and 18, thus making two scales. These yielded a reliability coefficient for the ten items of .88, which is quite high for such a small number of elements.†

The validity of the technique is shown in two ways. In the first place, the six teachers were asked to give their votes independently of the pupils. The six teachers named 51 of the 165 pupils, giving 31 positive votes and 20 negative votes. In only 5 cases were these votes in disagreement with those of the pupils.‡

THE CHECK LIST

The use of a word check list was suggested by the apparent success of this procedure in overcoming halo effects and the concentration of ratings at the upper end of a graphic scale. A full account of this device will be given in Volume III, as our chief interest in the technique was as a means of finding the subject's

* Various methods of scoring and scaling the items on the "Guess Who" test were tried out, which will be taken up in Volume III, when the significance of this device as part of a total criterion will be discussed.

† This does not necessarily imply either "halo" or the existence of a general trait of coöperativeness, as the various plus scores, *i.e.*, the various items for the same pupil, were given not by one person, but by several, and therefore represented different situations.

‡ The correlation between teachers' and pupils' votes was .80.

general reputation. Certain of the words, however, referred to the tendencies being studied here and so these were separately scored as a measure of reputation for helpfulness and coöperation.

There were two forms of the Check List, each giving roughly the opposites of the other, with 80 words in each list. The two forms were checked on different occasions, and each child was checked by two teachers where possible. The directions for checking and the words selected from the two forms as indicative of selfish or unselfish behavior are as follows:

Name of Pupil_____ No._____
Name of Person Checking_____
Contact with Child_____

Directions for Use of Check List

First read slowly all the words printed on this sheet. They represent various phases or types of character. When you feel thoroughly familiar with the words used, look at the name written at the top. Think a moment about this person and your various experiences of him or her. Then, *with this person constantly in mind,* read each of the words of the check list and, if it can honestly be applied to the person of whom you are thinking, put a check mark in the little box at the left of the word. If not, pass on quickly to the next word and consider whether you should check it, etc. Consider each word; *check as many as you wish to. Check none at all if none apply.* Do not put down any check unless you are fairly sure of your judgment. Now read these directions again before starting to check.

☐ Coöperative	☐ Compassionate	☐ Uncoöperative
☐ Obliging	☐ Public-spirited	☐ Selfish
☐ Thoughtful	☐ Solicitous	☐ Stingy
☐ Social-minded	☐ Charitable	☐ Quarrelsome
☐ Kind		☐ Cruel
☐ Considerate		☐ Inconsiderate
☐ Generous		☐ A bully
☐ Sociable		☐ Callous
☐ Unselfish		☐ Brutal
☐ Merciful		☐ A shirker
☐ Humane		☐ Greedy

The first and second columns show the positive words selected and the third gives the negative words.

In scoring the words indicative of reputation for service, each form was taken as a unit.* The number of negative words was subtracted from the number of positive words; and if the result was positive, the form was scored + 1 and if negative, − 1. If the plus and minus scores balanced, or if no service words were checked, the score was zero. As a matter of fact, in only two or three instances did any teacher check both plus and minus words for any one child, so that the zero scores for each form were only for the cases for which the teacher found no appropriate words to check: she either did not know the subject well or the subject was colorless.

Since there were usually two teachers' records and two forms for each teacher, the scores thus ranged from + 4 to − 4. When the teachers disagreed, one marking both forms plus and the other marking both forms minus, the resulting zero was called 0 −, the 0 being reserved for those receiving no checks at all. In cases where only one teacher's records were available and where two teachers marked only one form, the score was doubled, the zeros remaining zero or neutral.

This rough procedure, although it ignores some real differences among children above the average, is nevertheless quite reliable.†

---

* So as to even up the variation among teachers in the *number* of words checked.

† The curve of scores is J-shaped. A more refined method of scoring used for the entire check list will be reported in Volume III. A fourfold table, placing the zeros, of which there were only a few, with the plus scores, yields an r of .559. When the zeros are counted with the minus scores, the r is .402. When the zeros are omitted, the r rises to .738. These correlations are r_{AB} and in each case represent what is common to both forms when used by the same teacher. As we had the scores on both forms from each of two teachers for each pupil, the reliability of all four forms combined is figured by the following formula:

$$\text{Check List Reliability} = \frac{4\,r_{AB} + r_{(A_1+B_1)(A_2+B_2)}}{2 + 2\,r_{AB} + r_{(A_1+B_1)(A_2+B_2)}},$$

where the subscripts refer to the two teachers and the following assumptions hold:

THE CONDUCT RECORD

Here again we used only a portion of the rating device which was prepared as a measure of reputation in general, but which contained items bearing on the tendency to be helpful or the reverse. The scheme required the rating of the whole class at one time on a series of conduct tendencies variously classified as persistence, honesty, sense of justice, and the like. The following two items separately scored were used as a measure of reputation for service:

Coöperation:

(*a*) Works with others if asked to do so (+ 1)
(*b*) Works better alone; cannot get along well with others (− 1)
(*c*) Works well and gladly with others (+ 2)
(*d*) Indifferent as to whether or not he works with others (0)
(*e*) Usually antagonistic or obstructive to joint effort (− 3)

Selfishness:

(*a*) Self always first (− 2)
(*b*) Others first only under pressure (− 1)
(*c*) Others first if to his own advantage (0)
(*d*) Thinks of others first; little thought of self (+ 2)
(*e*) Self and others on equal basis (+ 1)

It may be seen that the types of conduct referred to under each item are roughly scaled. These scale values are printed after each descriptive paragraph. The coöperation scale is + 2, + 1,

---

1. That r_{AB} in the numerator is the average of the following:

$$r_{(A_1\mathrm{I})(A_1\mathrm{II})},$$
$$r_{(B_1\mathrm{I})(B_1\mathrm{II})},$$

etc. for all combinations of A_1I, A_1II, A_2I, A_2II, B_1I, etc., I and II representing the first and second applications of the Check List.

2. That $r_{(A_1+B_1)(A_2+B_2)}$ in the numerator $= r_{(A_1\mathrm{I}+B_1\mathrm{I})(A_2\mathrm{II}+B_2\mathrm{II})}$.

3. That r_{AB} in the denominator is $r_{(A_1\mathrm{I})(B_1\mathrm{I})}$ and that $r_{(A_1\mathrm{I})(B_1\mathrm{I})} = r_{(A_1\mathrm{II})(B_1\mathrm{II})} = r_{(A_2\mathrm{II})(B_2\mathrm{II})}$.

4. That $r_{(A_1+B_1)(A_2+B_2)}$ in the denominator $=$

$$r_{(A_1\mathrm{I}+B_1\mathrm{I})(A_2\mathrm{I}+B_2\mathrm{I})} = r_{(A_1\mathrm{II}+B_1\mathrm{II})(A_2\mathrm{II}+B_2\mathrm{II})}.$$

Now $r_{(A_1+B_1)(A_2+B_2)} = .55$. Taking as r_{AB} the average of .559 and .402 quoted on page 93, or .48, by this formula the reliability of the sum of the four forms becomes .704.

0, − 1, − 3. The selfishness scale is + 2, + 1, 0, − 1, − 2. Paragraph (*e*) under *Coöperation* is scored − 3 instead of − 2 since there were relatively few pupils who were checked as belonging to this category. The total criterion scores thus range from + 4 to − 5. It was impossible to secure second ratings on these items, so that no reliability figures are available for this device.

TOTAL REPUTATION FOR SERVICE

We have presented five different ways of finding out the reputation of the pupils in the matter of helpfulness and coöperation. As we secured at least three sets of opinions for each child, we need to form some idea of the accuracy and adequacy of the total reputation embodied in the sum of these three or four scores. This is estimated on the basis of the agreement among these various opinions and by the agreement of each with the sum of the remainder. In the case of population P, we used the Service Record, the Portraits, and the " Guess Who " scores. For populations X, Y, and Z, we had " Guess Who," Check List, and Conduct Record. We shall report the facts separately, therefore, for the populations having separate combinations of reputation scores.

In the case of population P, we made due allowance for the fact that one point on the Portraits meant far more than one point on the " Guess Who " or Service Record, and also for the fact that the " Guess Who " scores represented the opinion of a whole class as against the Portraits, which were made out only by the teacher. After weighting the scores, the relative significance of these three factors, " Guess Who," Portraits, and Service Records, was in the proportion of 12 : 7 : 9. The reputation score had a reliability of .91 and would probably have correlated with the children's total reputation secured by a large number of devices of this sort to the extent of .82.

For populations X, Y, and Z, we used three devices, giving scores ranging from + 20 to − 20 for the " Guess Who," + 4 to − 4 for the Check List, and + 4 to − 5 for the Conduct Record. In this total the Check List and Conduct Record were given equal weight

and the " Guess Who " twice as much as the two combined. This yielded a reliability coefficient of .88 for the combination and a predicted correlation with a theoretical total reputation of .77.*

* The basic figures for the correlations reported are given in Table XVII.

TABLE XVII

INTERCORRELATIONS OF MEASURES OF REPUTATION FOR SERVICE

TEST	POPULATION P		POPULATION XYZ	
	Service Record	Portraits	Check List	Conduct Record
"Guess Who" with .	.22	.61	.32	.32
Service Record with .		.41		
Check List with . .				.35

From this table we find an average inter-*r* for population P of .41, which yields a theoretical validity of .82 by Formula (13) (*Studies in Deceit*); an average inter-*r* for population XYZ of .33 and a theoretical validity of .77. That is, we should need 10 more similar (having the same average inter-*r*) records of reputation for P and 15 more for XYZ to achieve in each case a theoretical validity of .95.

If we take the sum of the two records from teachers as one unit and the records from pupils as another, the *r* between the two is .39, which gives a theoretical validity of .75. On this basis 12 more similar records would be needed for a theoretical validity of .95.

The reliability of the combined reputation score representing how well it would correlate with itself if *repeated* is .91 for population P and .88 for population XYZ, using "Guess Who," Check List, and Conduct Record. This reliability is figured by the formula

$$\text{Reliability} = \frac{\Sigma R + a(a-1)r}{a + a(a-1)r},$$

where R = the average reliability of the separate scores, a = the number of tests, r in the numerator = the average of the cross correlations, and r in the denominator = the average of the inter-r's on the assumption that this will be the same on the second occasion as on the first. We have also taken the cross r's as equal to the inter-r's and have estimated the reliability of the Conduct Record as .80.

The correlation of each of the devices used in population P with the sum of the other two is shown in Table XVIII.

TABLE XVIII

CORRELATION OF EACH OF THE SERVICE REPUTATION SCORES WITH THE SUM OF THE OTHER TWO — POPULATION P

DEVICE	r WITH REMAINING TWO
"Guess Who"	.440
Portraits	.560
Service Record	.300

CHAPTER V

THE VALUE OF THE SERVICE TESTS

NOTE: The general reader may prefer to turn at once to the concluding section of this chapter, on page 117.

DESCRIPTIVE RÉSUMÉ

In estimating the value of a test, one has in mind both scientific and practical considerations. From a practical standpoint, a test must be inexpensive and easy to administer. From the scientific standpoint, it must have satisfactory reliability and validity. The practical aspects of the service tests may readily be seen from the description given of them in Chapter III. This chapter deals with these two questions: Will they yield similar results if repeated? Do they measure a group of factors which may properly be classified as belonging together under the designation "service"?

Five tests were used in all, three of which involved the child's relation to his class group and two of which were of the more restricted "charity" type of behavior. The first three are the Free Choice test, the Money Vote test, and the Efficiency Coöperation test. The other two are the Kits test and the Envelopes test. The Free Choice test as used in population P * gave the pupil the chance to work for himself or his class in a spelling contest. In populations X, Y, and Z this was modified so as to allow the subject seven chances to work either for himself or his class in a contest in rapid additions. The Money Vote test required the pupils to decide how the money that might be won by the class in a contest would be spent. The Efficiency Coöperation test measured the relative amount of work done for self and for class in a rapid addi-

* Called there the Self-or-Class test.

tions test. The Kits test gave the children a chance to give away part or all of the articles just received as a gift, and the Envelopes test afforded an opportunity for them to find and bring in pictures, jokes, and stories for hospital children.

It may properly be questioned whether these two groups of tests belong together at all, in view of the introduction of the factor of loyalty in the class tests. Although the statistical values of the tests* do not provide evidence for thinking that this is the case, we have nevertheless endeavored to report the facts for the two types separately where possible.

In working through the discussions of this and subsequent chapters, the method of scoring the several tests should be borne in mind. In the Free Choice test as used in populations X, Y, and Z the maximum score is 7, as there were seven opportunities to prefer the class to self. The minimum score is zero, where all seven choices were for self. In the Money Vote test, the maximum score is for choosing to give the prize money for the most unselfish purposes, the next highest for school needs, the next for class needs. The lowest score is for a purely selfish choice. In the Efficiency Coöperation test, the highest scores go to the pupils whose work shows the greatest relative amount of work done for the class as compared with work done for self. The Kits scores range from a high score for those who gave away the most objects and the most cherished objects to zero, which would represent keeping the entire kit for oneself. Likewise in the Envelopes test, the highest scores were obtained by those who did the most work, with zero for those who did nothing. These were all common-sense procedures and were not determined by reference to any independent measure of coöperativeness.†

---

* The intercorrelations reported later in the chapter and the correlation of each test with the total criterion indicate as close a resemblance between the two types as within either type.

† The complicated procedures for weighting scores by reference to outside criteria were not used for two reasons: first, we wished to keep the test scores as "pure" as possible — *i.e.*, uninfluenced by the validation material; second, we had found by experience that the improvement in scoring resulting from elaborate statistical procedures was not great enough to warrant the expense.

THE RELIABILITY OF THE SERVICE TESTS

As our experimental work in population P was undertaken primarily to smooth out administrative problems and for the sake of validation, no direct facts concerning the reliability of the tests as given there are available. For populations X, Y, and Z, however, representing a wide range of social and intellectual level in two communities, we can estimate the extent to which the pupils would exhibit the same behavior under similar conditions or with the same or similar tests. Our estimate is not so satisfactory as it would have been if we could have repeated our tests after an interval of a few weeks. This was impossible except in the case of the Free Choice test. We have had to depend instead on such evidence as we could gather by a study of the test scores themselves. The facts will be reported test for test.

The Free Choice Test. This test was repeated after one week on a population of 120 pupils. The resulting self-correlation is .88, which means that the children were fairly consistent from one week to the next in the extent to which they sacrificed themselves in the interest of the class.

The Efficiency Coöperation Test. To find the reliability of this test, we divided the twenty-four units into two parts, the odd items in one part and the even in the other, as though the test had been in two forms. These two parts, consisting of twelve units each, gave a correlation of .82, which yields a reliability coefficient of .90 for the entire set of twenty-four items. Here again, one may feel reasonably sure that the behavior on this test is consistent.

The Money Vote, Kits, and Envelopes. No way was found for splitting these tests into equivalent halves, and they are not sufficiently alike to be regarded as parts of one test.*

---

* The Money Vote test was repeated in two of the population P classes. On the first occasion, the prize money was actually in hand. On the second occasion, it was only a possibility. In one class (a sixth grade) there was discussion of the first vote, but not in the other (a fourth grade). The sixth grade self r was .51, and the fourth grade, .65, in spite of the change in the situation.

The Total Service Score. As described in Chapter III, a total service score was found by combining the scores of the separate tests. As we are interested in measuring as large an area of coöperative behavior as possible, we need to know the reliability of the sum of all our tests combined. As we could not repeat the entire set and as we have reliability figures for only two of the separate tests, the best we could do was to combine half our material for one subtotal and the other half for another subtotal and to correlate the two. As we have already noted, our tests are not sufficiently alike to give us by this procedure as high reliability as would result from repetition, so that this result must be regarded as representing the minimum reliability.

For one sub-battery, which we shall call A, we took one-half the Efficiency Coöperation test, the Free Choice test, and the Kits. For the other sub-battery, B, we took the other half of the Efficiency Coöperation test, the Money Vote test, and the Envelopes. The correlation between the two sub-batteries is .64, which gives a reliability coefficient for the total of .78. We feel safe in estimating a reliability of not less than .80 for the total service score.

THE VALIDITY OF THE SERVICE TESTS

In our first volume, *Studies in Deceit*, we discussed the theory of the validation of conduct tests. By validity we mean the capacity of a test to measure what it claims to measure. This may be found by showing the amount of agreement between test scores and some other independent measures of what the test scores are supposed to represent, provided reliable sources of independent information are available. This we have called empirical validity. As noted in the previous chapter, however, there is a strong tendency among test builders to place greater emphasis on such matters as the internal consistency of a test and its capacity to represent the results of a very large number of similar tests. This we have called theoretical validity.

As our efforts to secure data concerning the children's reputation for service proved reasonably successful, we have thought it

wise to report the correspondence between reputation and behavior on our tests, but in forming our opinion of the real value of the tests we have depended chiefly on evidence of their reliability and theoretical validity.

CONDUCT AND REPUTATION

Our independent measures of the children's reputation for service were discussed in Chapter IV. This is essentially as objective a fact as the test scores themselves, although neither the test scores nor the reputation scores are identical with the inner tendencies of the individuals who are being tested. The total reputation scores run from a minus value through zero to a plus value. When a pupil is rated a low minus, it means that he has the reputation of being selfish, uncoöperative, and uncharitable. When he has a high plus score, we know that he has the reputation of being helpful, coöperative, and charitable. A zero or low plus score means that he is neither one thing nor the other, or sometimes one and sometimes the other.

Population P. The reputation measures for school P consist of the teachers' records of service rendered during the year in class and school projects, the " Guess Who " scores, and the Portraits. Taking the population as a whole, the separate service tests correlate rather low with this reputation score; and although, when combined with one another, they give somewhat better results, still there is not at once apparent any such agreement between tests and reputation as might perhaps be expected.*

More detailed examination of the facts, however, reveals some interesting things which are not inconsistent with the low correlation coefficients. For example, in the Kits test all those who were rated as below zero in reputation, who were regarded as definitely selfish and noncoöperative, scored below the mean on the tests. Furthermore, there is a little group scoring very high on the Kits all of whom rated above zero on the reputation scale. The same thing is true of the other tests.

* The *r*'s are shown in Table XIX, page 103.

Inasmuch as the tests taken singly would distinguish completely between those who had the reputation of being very selfish and those who had the reputation of being very coöperative, we felt inclined to go forward with the experiment and use our techniques, with what improvements we could command, in our major populations, X, Y, and Z.*

Populations X, Y, and Z. As in the case of population P, the separate tests of charity and coöperation do not agree well with the reputation the pupils have for this sort of behavior.† Here

TABLE XIX

CORRELATIONS BETWEEN SERVICE TESTS AND THE TOTAL REPUTATION FOR SERVICE — POPULATION P

SINGLE TESTS	r	COMBINED TESTS	r
Money Vote	.13	Kits and Envelopes	.32
Self-or-Class	.19	Kits, Envelopes, and Self-or-Class	.30
Kits	.32	Kits and Self-or-Class	.24
Envelopes	.30	Kits, Envelopes, and Vote	.31
		Kits and Vote	.30
		Envelopes, Self-or-Class, and Vote	.30
		Envelopes and Vote	.51
		Envelopes and Self-or-Class	.31
		Self-or-Class and Vote	.39
		Kits, Envelopes, Self-or-Class, and Vote	.31

* The right side of Table XIX is more encouraging. As tests are combined with one another, the r's with reputation rise, as is to be expected when the inter-r's are low.

† The facts are shown in Table XX, page 104.

A factor which may account in part for the relatively low r's of Table XIX is to be found in the nature of the "Guess Who" test. In this technique, it will be recalled, each child had a chance to give his opinion of every other child. There are some interesting sex differences in the way this opportunity was used. In X, Y, and Z, we found that in three classes the girls had rated the boys very low and themselves very high, although the boys had not shown any such sex prejudice. When we eliminate these three groups of boys and also two groups of girls whose teachers had given them abnormally high ratings as compared with the pupils' opinions and also with the fact that they did nothing on the

again, on account of the nature of the scores, the raw correlation coefficients do not tell the whole story. Let us look at the way in which the total reputation scores are distributed in Figure 11.

The scores run from − 40 to + 44, with 54% of the cases between 0 and + 9. The upper 23% have scores from 10 all the way to + 44, and the lower 23% run from 0 all the way to − 40. The

---

Envelopes test, the correlation of .52 between the total service reputation and test rises to .61, which is probably nearer the truth.

TABLE XX

Correlation of Service Tests with the Service Reputation — Population XYZ

	Check List	Conduct Rating	"Guess Who"	Total Reputation
Free Choice	.13	.18	.24	.27
Efficiency Coöperation	.25	.28	.37	.30
Money Vote	.03	.12	.19	.20
Kits	.15	.12	.25	.30
Envelopes	.44	.18	.15	.32
Total score				.52 (.61)

When the sexes are handled separately, the r's for the three populations are as shown in Table XXI.

TABLE XXI

Sex Differences in r's between Service and Reputation

	X	Y	Z	All
Boys	.66	.51	.58	.56
Girls	.41	.32	.57	.47

With the exception of the Z school, the correlations are markedly higher for boys than for girls. This would seem to indicate a greater degree of prejudice in favor of the girls, who are actually rated far higher than the boys — a prejudice which the tests justify largely, though not wholly; for although the girls are usually more helpful than the boys, they are not more coöperative on the Efficiency Coöperation test, but less.

narrow section of the scale, from 0 to + 9, really represents a large number of doubtful cases, as such cases are always thrown into

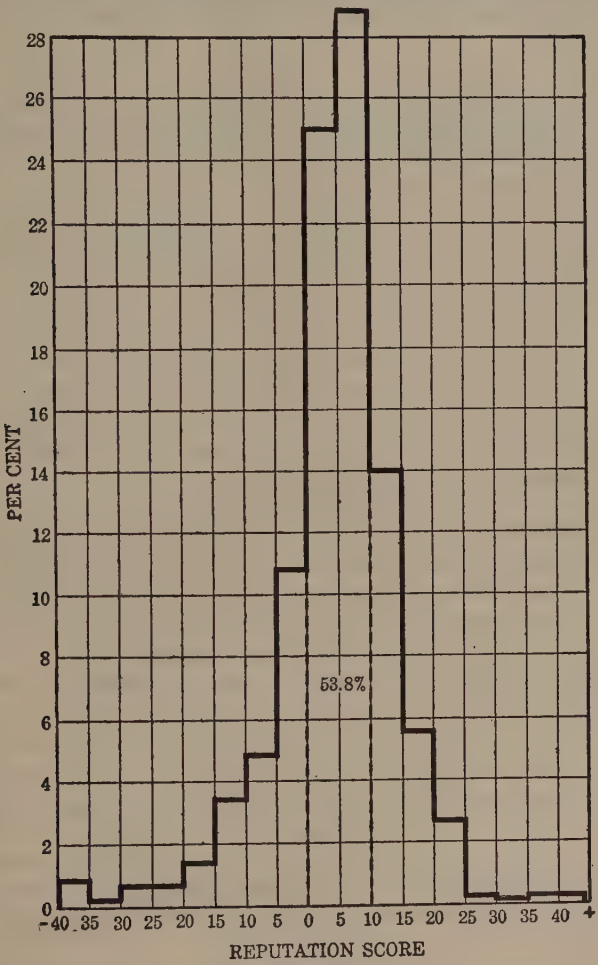


FIGURE 11

PERCENTAGE DISTRIBUTION OF THE SERVICE REPUTATION SCORES — ALL POPULATIONS COMBINED

N = 800

the middle. Considering only those who are sufficiently well known or whose tendencies are sufficiently developed to make an impression on their friends and teachers, we find the following facts for the several tests:

1. The Free Choice Test. Of the 165 who were rated as below zero in service tendencies, 40% failed to give any of their work to the class. Of the 177 who were rated above 9 on the reputation scale, about 23% failed to give anything to the class.

2. The Efficiency Coöperation Test. This does better than the Free Choice test. There are 25 cases which stand very high on this test. Of these only one was rated below zero in reputation. On the other hand, there were 35 pupils with very low coöperation scores, that is, who worked much harder for themselves than for their class, and of these there is only one who was ranked above 9, or as being definitely coöperative.

Again, of all those (261) who were above 7 on the reputation scale, only two showed distinctly selfish scores on the test. Of the 165 who had minus scores in reputation, there were only 13 whose scores showed that they worked harder for the class than for themselves.

Evidently this test does differentiate between those who have the reputation of being definitely coöperative and definitely noncoöperative, but it does not distinguish among those who are in the middle of the reputation scale. For about half the cases the correlation between test scores and reputation scores is high, but for those who are not so clearly distinguished in reputation the correlation with test scores is zero.

3. The Money Vote Test. Of the 179 who were rated below zero, 70% voted selfishly. Of the 170 rated above 9, 40% voted selfishly and only one received a zero score.

4. The Kits Test. Only 35% of the 160 receiving negative reputation scores contributed anything at all and 65% kept everything. But of the upper quarter of the reputation scale 60% gave something and 40% kept everything. Eleven out of the twelve rated highest gave something, and only four out of the twelve rated lowest gave anything. The 32 who gave away most had a median rating of 11 while those who gave nothing had a median rating of 2.

5. The Envelopes Test. Similar facts could be reported for this test. Of the 176 who had reputation scores below zero, 75% failed to do anything on this test. They brought nothing at all in the way of stories, jokes, and pictures for the hospital children. On the other hand, 48% of those

who were in the upper quarter of the reputation scale failed to do anything on the test. At the very extreme there were 36 pupils rated below − 14 (the lowest 5%). Of these only 11% did anything on the test. But 60% of those who were among the top 5% in reputation contributed something.

We may conclude that the service tests, when taken separately, distinguish fairly well between the top and bottom quartiles of the reputation scale, but that they will not differentiate among the middle 50%, who fall between the limits of 0 and + 9 on the scale.*

The Total Service Score. The correlation between the total service score and the total reputation score is .61.† Although this seems low, it should be borne in mind that the correlations between test scores and ratings for intelligence seldom run higher than .50. The entire correlation table is shown in Figure 12 and bears out what we have been saying about the separate tests. The heavy lines show that all those who score above 130 on the coöperation tests are rated above zero in reputation — that is, they are regarded by their teachers and fellow pupils as coöperative. On the other hand, all but one of those scoring below 95 on the service tests are rated below zero, or as definitely uncoöperative and selfish. There are 51 cases in the upper corner and 31 in the lower. These 82 cases, or 11% of the total, are completely differentiated by the tests, but the differential power of the tests rapidly diminishes as the center of the table is approached.

If we take quartiles, as we did for the separate tests, we find from the table that the upper quartile, or top 25% of the cases, includes those achieving a score greater than 119. Of these 94% were rated as being more coöperative than uncoöperative. The lower quartile includes those scoring less than 105. Only 64%

---

* We recognize that statements about extreme cases in scattergrams should be made cautiously because of the fact that many of these extremes may be due to chance. The chances of being able to make similar statements about similar scattergrams derived from fresh populations are somewhat more hazardous than statements about coefficients of correlation. But we have here at least three rather distinct populations, and nearly all of the foregoing statements will apply equally well to all.

† For correction for attenuation, see the discussion of "true validity," following.

of these were rated definitely minus. If a pupil has a service score of 120 or better, the chances are nine out of ten or better that he has a reputation for coöperation and helpfulness. If his score falls below 95, it is almost certain that he will have a reputation for selfishness and lack of coöperation.*

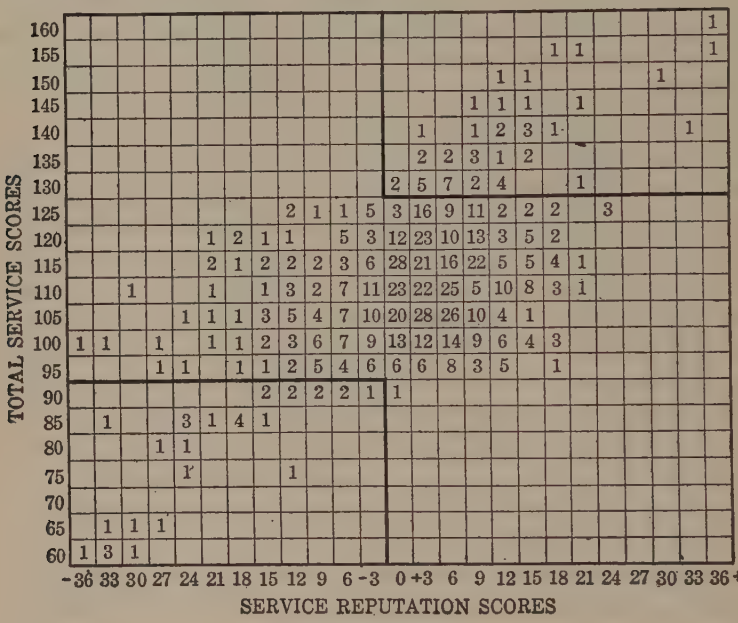


FIGURE 12

CORRELATION TABLE OF SERVICE SCORES AND REPUTATION SCORES — ALL POPULATIONS COMBINED

We showed by a graph (Figure 11) that about half the cases were rated between 0 and + 9 on the total reputation scale, which gives a very high peak on the diagram. The distribution of the total test score is very similar, as shown by Figure 13. Here we have approximately 50% of the cases scoring between 105 and 120,

* As all three of our population groups are represented in both the upper and lower quartiles, these results are not vitiated by errors of sampling.

which is only 15% of the entire range. In a normal frequency distribution, the middle 50% of the cases cover 23% of the range. In the case of Figure 11 it may be seen that the middle 54% of the cases are bunched over about 12% of the range of reputation scores.

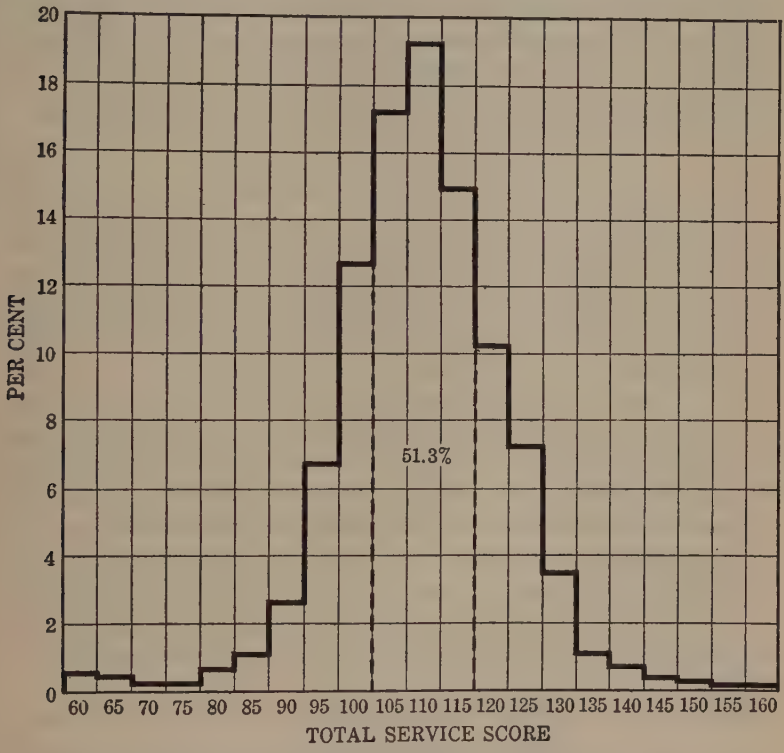


FIGURE 13

PERCENTAGE DISTRIBUTION OF TOTAL SERVICE SCORES — ALL POPULATIONS COMBINED

N = 800

Above and below this narrow central portion of the scale, the tests and the reputation agree fairly well. At the center they do not agree. This probably signifies that about half the children are definitely either coöperative or uncoöperative in enough situations to

have gained a corresponding reputation and that about half the children are so vacillating from situation to situation, or so little is known about them, that they cannot be reliably rated.* On any one test or scale they may be genuinely and accurately distributed. That is, real individual differences probably occur within this middle group. But these differences either are too small to be detected by their friends or by the tests, or else the pupils are so variable that they cannot be given a definite rating.

THEORETICAL VALIDITY

In *Studies in Deceit*, Book One, pages 140–142, and Book Two, pages 118–121, the idea of theoretical validity is discussed in some detail. It is assumed that each performance test is a sample of a large number of behaviors sufficiently alike to justify classifying them under some one name such as "deception" or "service." That is, these behaviors possess something in common, such as the acceptance of credit or advantage to which one is not entitled or giving aid to others. In the case of our service tests, we have five such samples. Furthermore, it is assumed that, in order to measure the entire range or area of behavior classified as honesty or coöperation or what not, it is necessary to measure enough samples of it to insure that the whole be adequately represented. Our problem is to determine the adequacy of our particular sample. This we may do by applying the formulas discussed in *Studies in Deceit*.†

We find in general that it would take a good many similar tests to cover the ground adequately, about seventeen, in fact, or twelve more than we now have. The five we do have, while serving very well to measure the particular types of charity and coöperation which they embody, would correlate with an infinite number of similar tests only about .745. This is a higher figure than we re-

* This peculiarity may account in part for the low inter-*r*'s among the tests.

† Two tables of intercorrelations are available for the determination of

ported for the correlation with reputation and is perhaps a better basis for determining the *behavior* significance of our battery, which purports to measure behavior and not motives.

---

the theoretical validity of our tests, one for population P and one for X, Y, and Z.

TABLE XXII

INTERCORRELATIONS OF PRELIMINARY SERVICE TESTS — POPULATION P

	2	3	4
1. Free Choice	.17	.12	.21
2. Money Vote		.14	.04
3. Kits			.40
4. Envelopes			

Average inter-*r*, .18

TABLE XXIII

INTERCORRELATIONS OF FINAL SERVICE TESTS — POPULATION XYZ

	2	3	4	5
1. Free Choice	.20 ± .02	.17 ± .02	.13 ± .03	.20 ± .02
2. Efficiency Coöperation		.27 ± .02	.32 ± .02	.21 ± .02
3. Money Vote			.27 ± .02	.12 ± .03
4. Kits				.12 ± .03
5. Envelopes				

Average inter-*r*, .201

Table XXII yields a predicted *r* with five other similar tests of .52 and a theoretical validity of .72. To give a validity of .90, fifteen more tests having an average inter-*r* of .18 would be needed.

Table XXIII yields a predicted *r* with five similar tests of .555 and a corresponding theoretical validity of .745. To give a validity of .90, twelve more tests having an average inter-*r* of .201 would be needed, and for a validity of .95, thirty-one more would be needed.

The correlations of each of the final tests with the total of the remaining four are as follows: Free Choice, .28; Efficiency Coöperation, .41; Money Vote, .33; Kits, .33; Envelopes, .25.

TRUE VALIDITY

We have not quoted the correspondence or apparent lack of correspondence between reputation and conduct as indicating the extent to which the service tests are valid measures of behavior of this type. Our chief reason for this unwillingness to depend on ratings has already been given: the ratings themselves are subject to error. Practically, therefore, the two existing sets of data, one from tests and the other from opinions, do not well agree. But we are able to make allowance for these errors in order to show the extent to which *true* measures of service and *accurate* opinions about such tendencies would agree. Accepting .88 as the reliability of the total reputation score and .80 as the reliability of the total service score, we find, by substitution in the appropriate formula,* that the r of .61 between service and reputation rises to .73, which corresponds closely to the theoretical validity of .745 just now reported.

It should be noted, however, that the reliability of .88 is a prediction of the agreement that would be found between the scores on our reputation tests and the scores on the same tests if they should be *repeated.* Teachers' ratings, however, probably include prejudice, and this prejudice operates to give a spuriously high correlation between one set of ratings and another. The same is true of the pupils' ratings. The combination of teachers' and pupils' opinions is therefore a better measure of reputation than either one alone. The prejudice (error) in one offsets that in the other.† But we may assume that they cover a certain common amount or area of experience correctly. The amount of this common opinion is expressed by the correlation between the teachers' and pupils' opinions, which is .39. If each were a *perfect* measure of the *entire* reputation of the pupils, the r would be + 1.00.

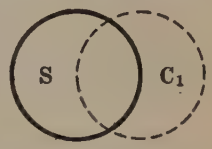
But each falls short of being a perfect measure of the entire reputation of the pupil in two ways. First, each contains a degree of

---

* See footnote on page 114.

† As these errors are probably correlated to only a slight extent.

prejudice or bias. It is well known that in judging the characters of others all judgments are more or less colored by personal likes and dislikes. Second, it is difficult to judge the entire field under consideration. Judgments are likely to be unbalanced by giving too much weight to relatively better known or possibly less important incidents.



This situation may be illustrated graphically as at the right: *

The solid line represents the entire area of situations and behavior known as service (S). The dotted circle represents the teachers' opinion (C) of the pupils' behavior in social situations. The opinions are "valid," that is, represent (apart from prejudice) the service tendencies, to the extent that the dotted circle overlaps the solid one.

Now if we introduce pupils' opinions of one another and assume that the agreement between them and teachers' opinion is due to the fact that they are both opinions about "service," and assume further that there is no agreement between teachers and pupils in such matters as personal prejudice toward any given pupil, then the situation may be shown in this way:

We now have represented an area that is common to conduct and reputation and also free from prejudice — for teachers and pupils at this point agree. This area is the small section of the figure common to the three circles, marked S. The dotted circles do not overlap outside the solid circle. Whatever agreement is here pictured between them has to do with service, nothing else.

Note also that neither of the dotted circles covers all the area of the solid one, and that each covers a part not covered by the other. But the two probably cover nearly all the area if we include what is not common to them. If ratings were available from parents and acquaintances, presumably the whole area would be covered

* Throughout this discussion these graphic illustrations are used merely to clarify the point. They are not meant to be geometric representations of fact.

eventually. The same reasoning may be applied to the tests. They cover part of the total area of service behavior but not all — and they also cover ground outside the service circle (that is, they measure something beside "service"). This is pictured as follows:

Now the validity of the tests (or of the ratings either) is the ground covered *inside* the service circle in proportion to that covered *outside*. If the test circle (T) exactly coincides with the service circle and if this coincidence is not a matter of chance, then the test has perfect validity. If we assume that the ground covered outside the circle has to do only with chance errors, then the extent of the ground covered inside is a measure of the validity of the test.

The proportion of the total service area covered by the test (ThV) is what we have been calling *theoretical* validity and is represented mathematically as the square root of the reliability of the test.

The next problem is to determine how much of the area of the service circle covered by the tests is also covered by the reputation. First we find the common ground between the tests and general reputation for service. This is represented by the r of .61. But part of this involves areas outside the circle, has to do with prejudice, etc. We wish to know what is common to test and reputation which is within the circle, that is, having to do with service only. This is represented by the correlation coefficient of .61 corrected for attenuation, which is 1.04.* This means that whatever is

---

* The purpose of the correction formula is to show the correlation between two sets of measures if each were free from errors. On the assumption that the errors of one set are uncorrelated with those of the other set, the formula becomes

$$r_{AB} = \frac{\sqrt{r_{A_1B_2}}\sqrt{r_{A_2B_1}}}{\sqrt{r_{A_1A_2}}\sqrt{r_{B_1B_2}}}.$$

In our case A_1 represents the tests and B_1 the opinions of teachers and pupils. $r_{A_1A_2}$ is the reliability of the five tests in the sense that the given five would correlate to this extent with another five having the same average intercorrelation, which in our case is .201. Hence $r_{A_1A_2} = .61$. The r between teachers' and pupils' ratings is .39. By the same prediction formula, these two would

common between teachers' ratings and pupils' ratings is also common between two series of similar tests.

Another way of stating the same fact is this: If we had a perfect measure of whatever the tests measure and also a perfect measure of whatever the reputation devices measure, these two would correlate 1.00.

We arrive now at a notion of the relation between empirical and theoretical validity. Theoretical validity is the extent to which the test actually covers the ground required (that is, the heavy circle); empirical validity is the overlap, within the circle, of the tests and the empirical criterion or reputation. If, however, the empirical criterion (reputation) coincides exactly with the circle, then empirical and theoretical validity are one and the same thing.*

The question now is to determine the amount of ground inside the service circle covered by ratings or by tests. We have five tests with an average inter-r of .201. Substitution in the proper

---

correlate with another two .56. In the numerator we use the r between the entire set of five tests and the combined judgments of teachers and pupils under each radical. Hence the formula, with substitutions made, is

$$\text{True validity} = r_{\text{tests reputation}} = \frac{.61}{\sqrt{.61}\sqrt{.56}} = 1.04.$$

* Mathematically the relation is this:

(1) The theoretical validity of the tests (ThVt) is the square root of the reliability of the test scores and (2) the theoretical validity of the reputation criterion (ThVc) is the square root of the reliability of the reputation scores, where reliability means the correlation of one series with another similar series of the same length and intercorrelation. Empirical validity (EmVt) is the correlation between test and reputation. The type of validity we have just been discussing, which we may call the true validity, or TrVt, is the ratio of empirical validity to the products of the two theoretical validities; that is:

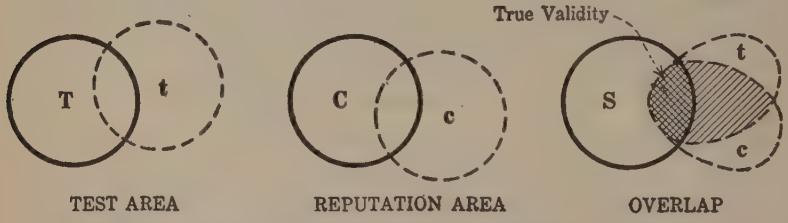
$$\begin{aligned} \text{ThVt} &= \sqrt{\text{Rt}} \\ \text{ThVc} &= \sqrt{\text{Rc}} \\ \text{EmVt} &= r_{tc} \\ \text{TrVt} &= \frac{r_{tc}}{\sqrt{\text{Rt}}\sqrt{\text{Rc}}} \end{aligned}$$

This may be seen from the usual formula for correction for attenuation, which assumes that the errors involved are uncorrelated.

formula* shows that we should need 36 similar tests to get an r of .95 with an infinite number of such tests. Hence our five tests cover about one-seventh of the ground. As for ratings, we have two sets — one from pupils and one from teachers, with an r of .39 between them. Substitution in the same formula shows that we should need 14 such measures to yield an r of .95 with an infinite number. Here again the two cover one-seventh of the ground. But the ground covered within the circle by the ratings is the same as that covered by the tests, as is shown by the true coefficient of validity of 1.00.

Our conclusion, then, is that the five tests of service which we now have are valid as far as they go but that they go only one-seventh of the way.

Supplementary Note. Strictly speaking, we should have begun by assuming two solid circles, one for reputation and one for tests. Let each circle be defined as " that area covered by an infinite number of similar measures." This definition avoids the assump-



tion that reputation and tests belong at all in the same circle. Thus in the figures above we have the test circle, one-seventh of which is covered by our present tests, and the reputation circle, for which we also have measures covering one-seventh of the total area.

When we bring these two figures together, they overlap, as shown at the right, to an extent represented graphically by the shaded area and mathematically by an r of .61. Within the solid circle the two dotted circles cover exactly the same ground. Now

* Kelley, T. L., *Statistical Method*, formula 159. See formula (14), *Studies in Deceit.*

the question is this: As we increase the number of measures so that the shaded portions grow larger for each circle, will the r between them always remain at 1.00? The answer is Yes, because the r of 1.00 was obtained on this assumption: it is the r between an infinite number of tests and an infinite number of ratings — in short, between the test circle and the reputation circle.

CONCLUSIONS

To sum up, our general conclusions concerning the value of these tests are as follows:

1. If the tests were repeated on the same pupils under similar conditions, the total scores of the two occasions would be similar enough to give a correlation higher than .80 between them. This means that whatever behaviors these tests measure they measure consistently.

2. The tests will differentiate in a very clean-cut manner between those pupils who are regarded by teachers and classmates as definitely helpful, coöperative, and charitable and those who are regarded as the opposite. They will not differentiate, however, among those of the middle ranges who are sometimes coöperative and sometimes not or who are only moderately coöperative.

3. The low inter-r's between them indicate that they have little in common. The directions given to the pupils, the conditions under which they worked, the scoring plans (all described in Chapter III) indicate that each test presents to the pupil a different kind of problem. In the Free Choice and Efficiency Coöperation tests the appeal is to class loyalty; in the Kits and Envelopes it is to charity or generosity. But the tests that appeal to class loyalty have a low r (.20) between them, which indicates that they also measure different things; in the Free Choice test the score is the number of decisions made to give one's work to the class; in the Efficiency test it is the total amount of work done for class as compared with work done for self. Evidently one is very different from the other. Also in the Kits and Envelopes we have a similar contrast. In the Envelopes the task was *doing* something for

others; in the Kits it was *giving away* something. Again from the low r (.12) between them, these are evidently two quite different kinds of charity.

4. In our tests, as in our measures of reputation, we have covered approximately one-seventh of the ground which a complete battery would cover; but so far as we have gone, the tests are valid in the sense that reputation and test behavior agree with regard to the field common to both.

PART II

Factors Associated with Service

CHAPTER VI

AGE, SCHOOL STATUS, AND INTELLIGENCE

The major interest of the Character Education Inquiry is to investigate the possibility of measuring character. We have been gradually building up a series of tests for the measurement of tendencies which have a common-sense relation to character as this term is generally understood. The value of the most favorable combination of these types of behavior as indicative of character will be discussed in Volume III. Meanwhile we have been learning a good deal about the particular tendencies dealt with by each of our several groups of tests. As these tendencies, namely, deceit, service, inhibition, and persistence, are all of considerable importance in human affairs, it has seemed wise to report such facts as we have been able to gather about each tendency and its causes and characteristics.

The data gathered by the Inquiry may be roughly divided into four parts: (1) the scores on the character tests, (2) validating materials in the form of teachers' ratings, pupils' opinions of one another, records of social functioning in home, school, church, and club, opinions of Sunday-school teachers and others; (3) scores on other tests, such as intelligence tests, home background tests, emotionality tests; also a wide variety of information on school status, such as grade, retardation, deportment; and (4) data on educational procedures in instances where we have been able to compare children who have been subjected to different educa-

tional plans. Data of types (1) and (2) were gathered for the purpose of test construction and validation. Data of type (3) were gathered because we could get them easily and readily without much additional expense, and from them certain causal relations may be determined. Data of type (4) were gathered in connection with *Studies in Deceit* and appear in Part III of Book One of that volume. Since we have not carried the present tests to the point of attempting to appraise educational procedures with them, few facts of this type are available in connection with our study of service.

The methods used in analyzing the collected facts are the methods used in our first volume. By taking a large number of cases and dealing with averages, standard deviations, correlation coefficients, and other statistical devices we hope to untangle, at least in part, the intricate and complicated web of causal factors that lies back of each type of conduct included in our exploratory research. These methods are described in *Studies in Deceit* and justification for using them is given. Since the report of the Inquiry is regarded as a unit, we shall not repeat here what was said in that volume.

AGE AND SERVICE

As children grow older they increase in height and weight and other bodily structures until a certain age is reached. The relations of many anatomical and physiological structures and functions to age have been scientifically determined and charted, so that we have age-height charts, age-weight charts, age-teeth charts, and so on. We also know that intelligence increases with age up to a certain age level, and we accordingly have mental age charts. Furthermore, school progress is charted by age, so that we know what age is normal for each school grade.

In a certain crude way behavior, attitude, and emotional expression are graded by age. There are types of conduct and attitude that are regarded as "childish" or as "manly" or "womanly." But the age scale for conduct and attitude is very crude indeed and is based on tradition, not science. One of the problems before us

is to find out the extent to which it is possible to chart conduct according to age. Do children become more charitable, coöperative, and helpful as they grow older? Are older children, who have had more social experience, more or less selfish than younger children? Questions such as these are difficult ones and we may hope to find only partial answers in such samples of conduct as we have taken.

First we look at the scattergrams* showing the relation between age and our total performance service score for the three major populations studied. Here we find one zero correlation and two slightly negative correlations. For populations X, Y, and Z they are −.203, .052, and −.158,† respectively, and −.18 for all combined.‡ This means that in population Y there are no significant differences between the age groups in the scores obtained in the service tests when all tests are combined, and that in populations X and Z the younger children tend to score a little higher than the older ones.

Taken at their face value, these figures are somewhat surprising, indicating as they do that in these populations the older children do no more, and indeed do a little less, than the younger ones when opportunities to be of service are presented. This confirms the tentative conclusions given in Chapter I as the result of our preliminary experiments with public school children. A more detailed analysis of this relation is shown in Figure 14, where we report

---

* A scattergram, the reader will recall, is a device for showing the amount of agreement there is between two sets of facts about the same individuals. These and other statistical terms used in this study are discussed in *Studies in Deceit*, Book One, Chapters V and VI. A scattergram is shown in this volume on page 108.

† Exclusive of grade 8, which had some irregularities in the scores. If these had been included, the r for Z would have been − .34.

‡ The partials, with intelligence constant, are only slightly less, the − .18 changing to − .15. That is, even though all children were of the same intelligence, the correlation with age would not be much greater than reported. Another way of eliminating the effect of retardation and acceleration is to omit from the scattergram all those much over or under age. We did this for population Y, and the r of .052 rose only to .159.

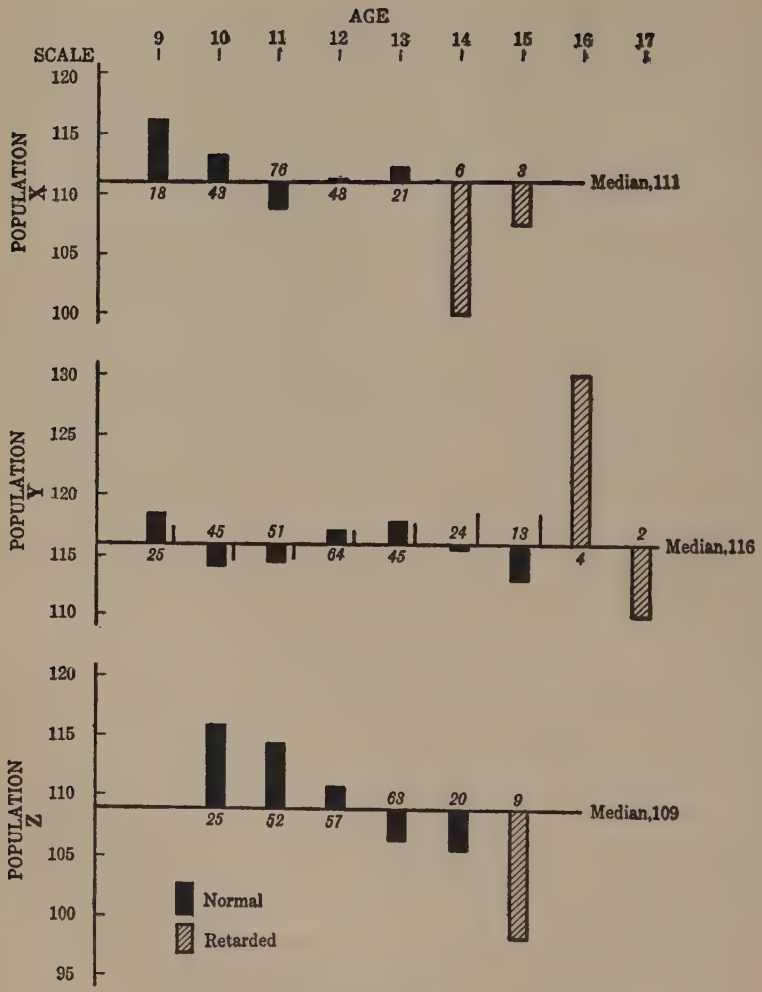


FIGURE 14

MEDIAN TOTAL SERVICE SCORES BY AGE COMPARED WITH THE MEDIAN FOR ALL — POPULATIONS X, Y, AND Z

The N for each age is shown in small figures at each bar.

the median test score for each age group in each population (Table I, Appendix C).

In this figure the few children who are distinctly older than others of the same school population are shown in shaded bars. With the exception of four sixteen-year-old children in population Y, all these retarded pupils are less helpful than the average of their own population. If they are all omitted from consideration, we note that in population X the scores decrease slightly as the ages increase. In Y they increase slightly with age. In Z they fall rather sharply as one goes up the age scale. The case for Y is still clearer if all pupils who are either much older or much younger than their classmates are omitted. The medians for the children who are left are shown on the figure by a narrow line at the right of each bar. The increase with age is considerable.

When this situation is analyzed further for each separate service test, we find that the scores remain about the same for each age on the Efficiency Coöperation test, the Free Choice, and the Money Vote, but drop with age on the Kits and Envelopes. These facts are summarized by the correlations in Table XXIV.

TABLE XXIV

CORRELATIONS BETWEEN AGE AND THE SEPARATE SERVICE TESTS

POPULATION	FREE CHOICE	EFFICIENCY COÖPERATION*	MONEY VOTE	KITS	ENVELOPES
X	.049	− .042	− .128	.091	− .423
Y	.044 †	.116 †	.052†	.000 †	− .125 †
Z	− .113	− .162	− .151	− .285	− .198
XYZ	.028	− .101	− .092	− .125	− .277

From the correlation table it appears that the younger children tend to respond more freely on the Envelopes test than the older ones. In population Z, which is the underprivileged community, the same is true for all the tests. A more detailed picture, in terms

* The biserial r's are about 50% larger.

† The corresponding r's, omitting accelerated and retarded cases, are, for Y, .057, −.128, −.025, .095, −.106.

of the *proportion* of each age group which helped at all on the Kits, Envelopes, and Free Choice tests is shown in Figure 15. Here again, it may be seen that in population Z fewer children coöper-

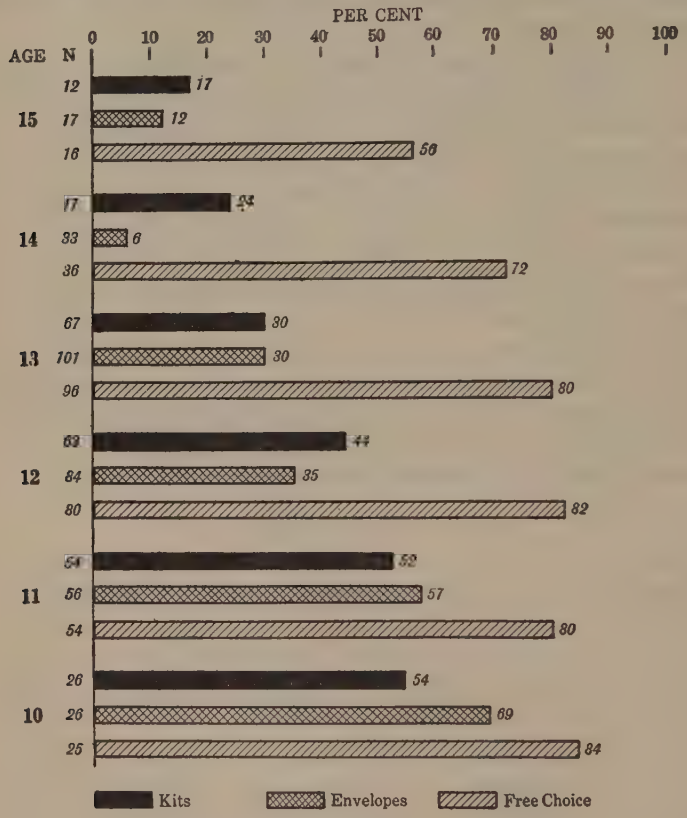


FIGURE 15

PER CENT AT EACH AGE COÖPERATING ON THREE SERVICE TESTS — POPULATION Z

ate. The same change with age is found in the privileged community, X, in the case of the Envelopes test. In population Y, which is a more normal community, no consistent differences

from age to age are found on the separate tests, although there is a tendency for the total score to increase slightly with age, as already pointed out. In population X, the percentages coöperating at ages 9 to 14 are, respectively, for the Kits test, 71, 68, 59, 55, 81, 33; for the Envelopes, 67, 48, 28, 26, 14, 0; for the Free Choice, 72, 53, 70, 77, 74, 67. In population Y, ages 9 to 15, the corresponding per cents are: Kits, 71, 68, 54, 58, 69, 67, 79; Envelopes, 62, 70, 60, 63, 57, 40, 60; Free Choice, 67, 64, 59, 80, 80, 81, 54.

Before final conclusions are drawn concerning the relation of age to the particular forms of helpful behavior with which we have been dealing, another set of important facts should be presented. A moment's reflection will lead the reader to see that "age" as such does not exist in a simple and uncomplicated biological form. School children are stratified artificially in grade groups and progress from year to year according to their ability to do school work rather than according to their physiological maturity. So far as school populations are concerned, therefore, grade is a more basic fact in determining the nature of experience than is age. Let us turn, then, to the consideration of school grade in relation to service.

GRADE AND SERVICE

The problem is to determine how service scores correlate on the whole with school grade. Since there are but four grades involved, we have computed the mean service score for each grade separately. These figures are given in Figure 16.*

In population X there are two schools, one of which contains grades 5 and 6 and the other 7 and 8. Consequently the grade differences for X are not as comparable as those of the other two populations, each of which is one school. The fact that the fifth and sixth grades in X score higher than the seventh and eighth may be due to grade or to school morale. Both schools, however, are in the same kind of environment. So it may be that children of the well-to-do classes grow less coöperative as they ascend the grades.

* The complete table from which this figure is drawn is Table II, in Appendix C.

The case of populations Y and Z is much clearer. In Y the total service scores definitely increase as the grade increases, but in Z the reverse is true. It is interesting to note that the fifth grades of both Y and Z have the same mean service score, but

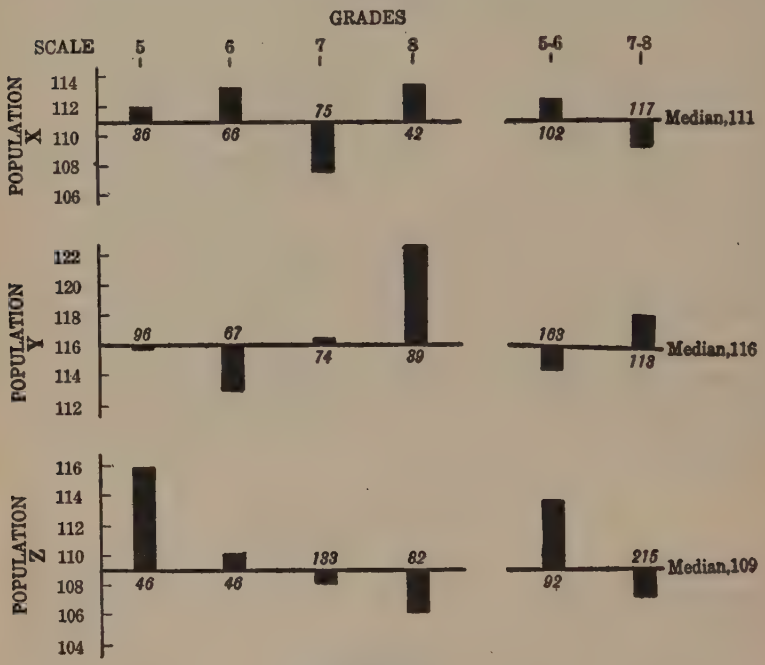


FIGURE 16

MEDIAN TOTAL SERVICE SCORES BY GRADE COMPARED WITH THE MEDIAN FOR ALL — POPULATIONS X, Y, AND Z

The N for each group is shown in small figures at each bar.

between the sixth grades there is a difference of 3 points, which is not significant; between the seventh grades the difference is 8 points, and between the eighth grades it is 16 points. When the fifth and sixth grades are combined, the difference is one point; when the seventh and eighth grades are combined in each population, the difference between Y, seventh and eighth, and Z, seventh

and eighth, is 11 points. These differences hold true for both sexes, which fact the chart does not show.

These facts correspond roughly with the correlations between age and service. The r was slightly positive in Y and negative in Z. There can be no doubt that the upper grades and to a certain extent the older groups in Y have a higher average service score than the younger groups and lower grades; in Z the reverse is true. A possible explanation may be found in the contrasting environment. At the level of the fifth grade both Y and Z are the same and in fact both ahead of X. It might seem that the longer the children remain in the Z type of environment, the less helpful and charitable they become; but the longer they remain in the Y type of environment, the more helpful and charitable they become. As we shall show in Chapter XI, however, this hypothesis is not tenable in this form.

Let us turn, then, to the consideration of the separate tests to see in what kinds of service this grade change is most marked.

In populations X and Y, only small and inconsistent grade differences occur in average scores on the Efficiency Coöperation test and Money Vote test. In Z, however, as may be seen from Figure 17, the upper grades fall off, particularly on the Efficiency Coöperation test. In per cent coöperating on the other three tests, however, some important grade differences are found. These tests represent the self-sacrificing type of behavior, as the pupil must choose between himself and others and between spending time on himself and on others. In Kits and Envelopes, the sacrifice is for unfortunate children. The percentage who gave away anything at all from their school kits tends to increase with grade in population X (53%, 60%, 62%, 74%), but not consistently in Y (68%, 53%, 57%, 88%); the seventh and eighth grades combined are superior to the fifth and sixth in both X and Y; but the drop in Z is startling. In this unfavored community the eighth-grade children did not have the Kits test, but Figure 18 shows that, while 65% of the fifth and sixth grades gave something, only 23% of the seventh grades gave anything. The Envelopes test (the cross-hatched bars) shows the same results in Z. In 5 and 6, 53% coöper-

ated, while in 7 and 8 only 27% contributed anything. In Y the upper grades have a corresponding record of 61% as against 57% for the lower, but in X the drop is from 49% for 5 and 6 to 16% for

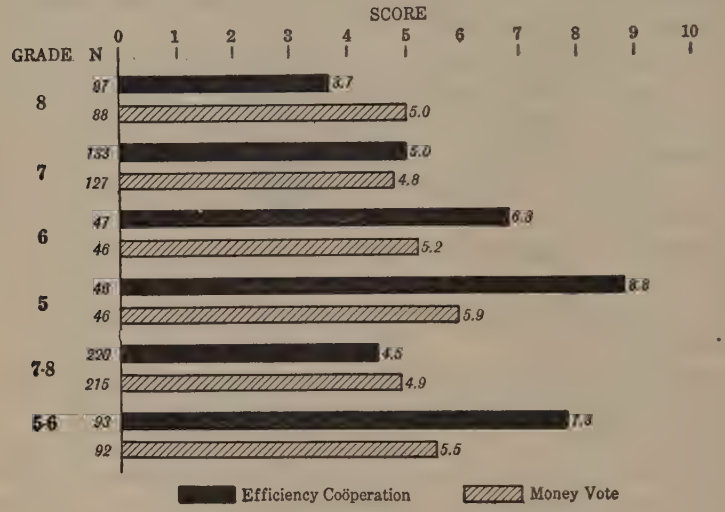


FIGURE 17

MEAN SCORES BY GRADE ON EFFICIENCY COÖPERATION* AND MONEY VOTE TESTS — POPULATION Z

7 and 8. As already stated, however, the case of X is not exactly comparable, since 5 and 6 are in one school and 7 and 8 in another.

In the Free Choice test, shown also for Z in Figure 18 (shaded bars), there are no appreciable grade differences. In X and Y

*The raw score on the Efficiency Coöperation test is the average difference between the examples done per minute for self and the examples done per minute for class. In this figure the raw score was subtracted from 10 for convenience of presentation. Thus the mean score of 6.8 represents a mean difference between the self and class scores of 3.2 examples. The Money Vote scores are a scale which ranges from 0 to 10, the higher scores representing the more charitable outlook. Complete data are presented in Tables III and IV of Appendix C.

the *upper* grades have the larger proportion of coöperators — 79% as against 63% for X and 85% as against 61% for Y.

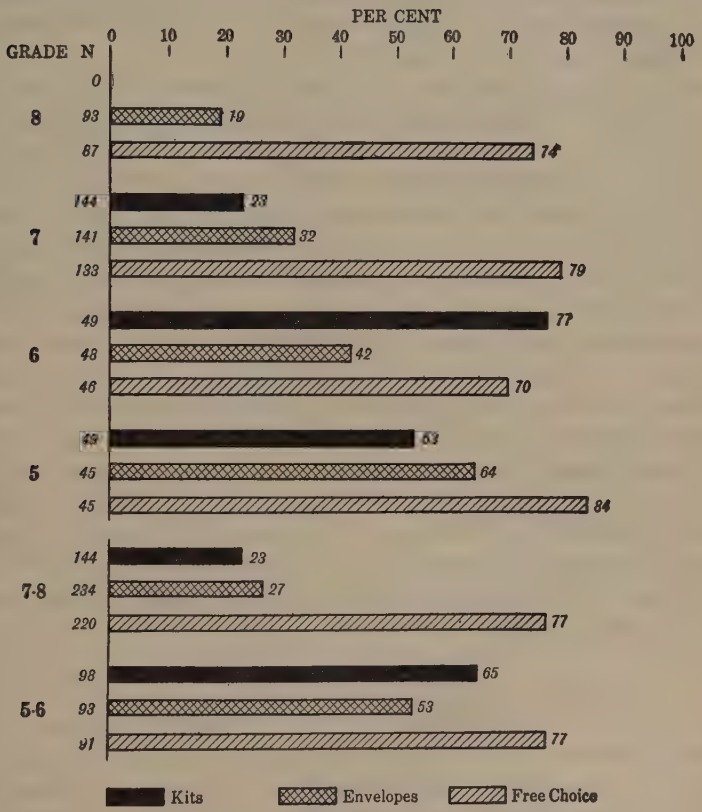


FIGURE 18

PER CENT OF EACH GRADE COÖPERATING ON THREE SERVICE TESTS — POPULATION Z

Table V, Appendix C, shows that the fifth grade, Z, not only has a very large percentage giving at least one exercise to class, but also has an exceptionally large average *amount* given. Why this one class should stand out on this test is not known.

To sum up, the lower grades in Z are superior to the higher in all tests except Free Choice. In population Y the upper grades are superior to the lower in all tests except the Efficiency Coöperation. In population X, grades 5 and 6 are superior in Envelopes and Money Vote but inferior in the other three.

Are these grade differences due to home environment or school environment? We have facts which reveal a general tendency for the older school children of our populations to come from the better homes, probably because children from poor homes tend to drop out of school early for economic reasons and also because they are likely to be on the whole less able. This might explain the fact that in Y the service scores increase with grade. But how about Z? Here they decrease with grade. The home environment hypothesis will not account for both sets of facts. A school environment or school morale hypothesis fits the case much better. In *Studies in Deceit* we found traces of cumulative school influence. This appears to be what is happening here. In the Z school, where the morale is not high, the influence of the school on group morale accumulates from year to year. It was noted that grade 5 in this school is rather high on all tests. Apparently the school influence begins to pull down the service scores about the sixth grade. This hypothesis is further supported by data on group morale which will be presented later.

RETARDATION AND ACCELERATION

It is a well-known fact that in any single school grade the children range in age over a period of three to seven years. Wide age ranges within a grade mean that some children are accelerated (*i.e.*, young for the grade) and others are retarded. It is also well known that retardation and acceleration sometimes create troublesome situations. A child who is too old for his grade is usually larger and stronger physically than the other children. He feels out of place and hence does not really function as a member of the group. Equally out of place socially are the very youngest children of the grade, who are usually small and relatively weak

physically. Hence retardation and acceleration are in themselves facts of great social importance. With the exception of the sixteen-year-olds of population Y, Figure 14 shows that all those who are over age for their respective schools score relatively low on the service tests. Evidently the problem of retardation and acceleration deserves our attention.

There are two ways of thinking of the grade status of a pupil. In studying the efficiency of school systems or the grade placement of school subjects, retardation can best be computed in terms of relation to some general age norm for a particular grade. The facts uppermost in mind have to do with intelligence, the ability to handle certain problems or texts, and the like, which have a fairly definite position in the course of study. We are dealing, however, with a different set of facts. It is the pupil's behavior in relation to his classmates, his teacher, and to persons outside the school in which we are interested rather than his competence to attain a certain standard ability in spelling or arithmetic. Hence the relation of a child to a theoretical grade norm is of far less significance than his status in his own particular classroom. If a whole group is retarded, the position of the members to one another may not be affected in the least by that fact. There would still be pupils in such a class who would seem relatively bright and advanced, for whom the psychology of acceleration would be the same as for those who would be classed as accelerated by the more usual definition. We are dealing with immediate social situations in the class group rather than with the relation of pupils to a course of study.

These two views of acceleration and retardation require two different methods of figuring the grade status of the pupils. We have used both and can report the facts in their relation both to the social experience of the pupils and to their academic status in the curriculum.

Scholastic Grade Status. In the United States it is customary for children to begin the first grade when they are six years old and entering the seventh year of their lives. If they progress a grade a year, they will be ten years old when entering the fifth grade. As

a matter of fact, however, after allowance is made for absence, for variations in grading, etc., the median age of children on most of the populations with which the Inquiry has dealt heretofore is nearer eleven than ten on entering grade 5. In the present populations, however, a child in the fifth grade and between ten years and ten years and eleven months old in November is normal for his grade,* and a year is added for each higher grade.

The percentage distribution of normal, accelerated, and retarded children for each grade of each school† shows that the X children

* We recognize that this is putting our norms almost a half year lower than we did in *Studies in Deceit* and almost a whole year lower than McCall's norms.

† The following tables show the grade status of each population. Table XXV gives the per cent falling at the standard normal age and at each year of acceleration or retardation.

TABLE XXV

PERCENTAGE DISTRIBUTION OF GRADE STATUS BY YEARS FOR EACH POPULATION AND GRADE

GRADE	POPULATION	PER CENT RETARDED				PER CENT NORMAL	PER CENT ACCELERATED	
		4 years	3 years	2 years	1 year	10 to 10:11	1 year	2 years
5	X			2	15	51	30	2
	Y	7	3	9	20	34	24	2
	Z	2	4	15	28	48	2	0
						11 to 11:11		
6	X	0	1	0	9	57	27	5
	Y	1	6	14	25	18	18	18
	Z	4	2	17	35	33	6	2
						12 to 12:11		
7	X	0	8	3	8	47	34	0
	Y	0	5	12	25	40	18	0
	Z	0	6	11	38	27	18	0
						13 to 13:11		
8	X	0	0	3	3	20	57	17
	Y	4	4	11	13	48	21	0
	Z	0	0	7	23	45	23	2

are more accelerated than retarded and the Y and Z children, particularly Z, more retarded than accelerated.

Service Scores of Well and Badly Graded Children. In Appendix C, Tables VI to VIII, all the facts bearing on scholastic status are brought together. There is no consistent tendency either in one population or in one grade for scholastic grade status to affect the service scores. Some extreme differences occur among the three groups, but sometimes it is the normal who are superior, sometimes the accelerated, and sometimes the retarded. In population Z, the retarded group, constituting the majority, are far less helpful than the others in all but the fifth grade.

We might expect that, if a difference in several tendencies were to occur at all between normal and retarded or accelerated pupils,

---

It can be seen from Table XXV that, if the norm had been moved up a half year, the distributions for Y and Z would not be so skewed. This would have made the grade norms the same as those found for *Studies in Deceit*. But Table XXVI shows that the present norms correspond to the modal year at each grade and move up one year at a time, beginning with 6 years and 2 months in grade 1. The use of medians instead of means would have increased each norm about a half year.

TABLE XXVI

PERCENTAGE DISTRIBUTION OF ALL CASES BY AGE AND GRADE

AGE	PER CENT OF GRADE			
	5	6	7	8
8	2			
9	21	3		
10	**41**	19		
11	22	**41**	21	3
12	8	23	**32**	29
13	3	8	30	**39**
14	3	3	11	19
15		2	6	6
16				2
17				1

Table XXVII summarizes Table XXV.

it would be most apt to show itself in such tests as set the class over against self as an object of interest. When all grades and populations are combined, the retarded group is found to be the least coöperative, the normal group the most coöperative, with the accelerated children coming second. This is usually true for both tests and for the four grades and the three populations. The differences are not great, but they suggest that the pupils who are most correctly graded are rather more inclined to feel the pressure of group loyalty than are those who are more or less out of step.

Social Grade Status. In order to study the social influence of retardation and acceleration, we first distributed the pupils of each class group of population Y by age in months, treating boys and girls separately. We then counted in from each end of the distribution in such a way as to block off children who were obviously

TABLE XXVII

PERCENTAGE DISTRIBUTIONS OF NORMAL, ACCELERATED, AND RETARDED GROUPS BY GRADE AND POPULATION

GRADE	POPULATION	PER CENT RETARDED	PER CENT NORMAL	PER CENT ACCELERATED
5	X	17	51	32
	Y	39	34	26
	Z	49	48	2
6	X	10	57	32
	Y	46	18	36
	Z	58	33	8
7	X	19	47	34
	Y	42	40	18
	Z	55	27	18
8	X	6	20	74
	Y	32	48	21
	Z	30	45	25
All	X	13	48	39
	Y	38	36	26
	Z	48	37	15

either much older or much younger than their classmates. We tried to leave about two-thirds* of each class in the normal group. We then combined all the "normal" children as thus empirically selected into one large group; and so with the "accelerated" and "retarded," keeping boys and girls separate. The service scores of each of these groups were then distributed with the results shown in Tables IX to XIV of Appendix C.

The accelerated boys and girls prove more helpful than either the retarded or normal, with the retarded considerably the least

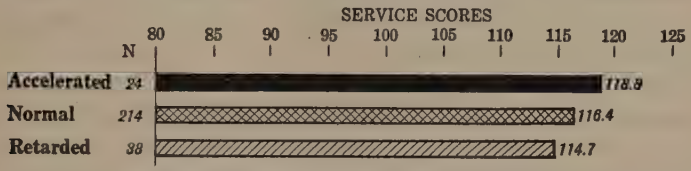


FIGURE 19

MEDIAN SERVICE SCORES OF NORMAL, ACCELERATED, AND RETARDED PUPILS OF POPULATION Y

coöperative. Figure 19 shows the results for the two sexes combined.

The small number of cases makes these differences somewhat unreliable,† and before placing great confidence in them corresponding figures for the separate tests should be consulted.

Social Grade Status and the Separate Service Tests. The scores on the two most distinctly coöperative tests should be compared with those of the Kits and Envelopes.‡ In per cent helping on the Kits, the normal boys (67%), curiously, are superior to both the retarded (54%) and accelerated (56%), with a considerable gap between normal and retarded. But the normal girls (61%) are inferior to both retarded (65%) and accelerated (85%), with

---

*That is, about one sigma above and below the average.

† The difference between the means of normal boys and retarded boys is 1.2 sigma; between normal and retarded girls, .8 sigma; and between the normal and retarded of both boys and girls, 1.5 sigma.

‡ See Tables XI, XIII, and XIV in Appendix C.

the accelerated far in advance of all. When boys and girls are combined, the order is accelerated (73%), normal (64%), retarded (59%).

On the Envelopes test the accelerated girls (85%) again do better than the normal, but here the retarded (64%) fall below the normal (73%), as in the case of the boys, where the proportions are 55% for accelerated, 48% for normal, and 35% for retarded. The order for both combined is accelerated (71%), normal (60%), retarded (46%).

The retarded boys and girls both vote more generously than the normal on the Money Vote test, and the order for both together is retarded, accelerated, normal.

So much for the charity tests. The coöperation tests may best be compared by reference to median scores. The boys show no effects of acceleration or retardation on the Efficiency Coöperation test although the retarded girls, scoring 6.9, are distinctly less coöperative than the normal and accelerated, who score respectively 8.8 and 9.2. The accelerated boys are decidedly more coöperative than normal boys on the Free Choice, however, giving away 3.5 units of work on the average, as against 2.6 for the normal; and the retarded are a little less coöperative than the normal, with a score of 2.3. On this test it is the retarded girls who do well; scoring 3.2, with the accelerated scoring 1.8, which is slightly below the normal score of 2.0.

It is somewhat surprising to find that the boys' behavior on the coöperation tests depends very little on the social experience of being retarded or accelerated, whereas the retarded girls are decidedly superior on the Free Choice and inferior on the Efficiency test, which requires endurance as well as generosity. Contrary to our expectation, the largest differences between the three groups appear on the Kits and Envelopes tests; but again it is the girls who are most affected, with the accelerated group superior on each test.

Social Grade Status and the Criterion. That the relationships among the well and badly graded children which we have just reported are valid and true is indicated by the fact that they agree with the opinions of teachers and pupils. These opinions, which

are summarized in our total service reputation score, are given in Table XV of Appendix C. Figure 20 summarizes the facts. As we shall point out in the next chapter, the sex differences are extreme; but with both boys and girls the accelerated are regarded as the most coöperative and helpful, then the normal, and lastly the retarded. Furthermore, these differences are greater for girls than for boys.

Retardation and Acceleration in Populations X and Z. The results just discussed were checked in a small way for four class-

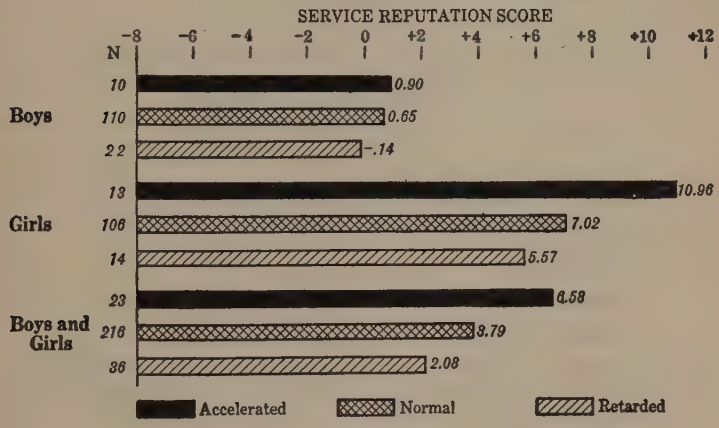


FIGURE 20

MEAN SERVICE REPUTATION SCORES OF ACCELERATED, NORMAL, AND RETARDED BOYS AND GIRLS — POPULATION Y

rooms in populations X and Z, where the distribution of ages was such as to permit a fairly clear-cut distinction between normal, accelerated, and retarded. The complete distribution of the median service scores for each age level in the four classes for the total and for the Free Choice test and the Efficiency Coöperation test will be found in Table XVI of Appendix C. In the fifth grade the retarded are usually the more coöperative, but in the three other classes the retarded fall far below the normal both in the total score and in the two coöperation tests. The extremely

accelerated also tend to fall off on the total scores, though but little on the coöperation tests; indeed, in some instances their scores exceed those of the normal group.

Summary of Effects of Retardation and Acceleration. Our study of the facts of social retardation and acceleration has included population Y with five tests and the total score and four classes in X and Z with two tests and the total score in each case, making 18 records in all. A rough assignment of the retarded, normal, and accelerated groups to first, second, or third position in helpfulness yields the following diagram:

NUMBER OF INSTANCES IN WHICH THE ACCELERATED, NORMAL, AND RETARDED GROUPS WERE FIRST, SECOND, OR THIRD IN HELPFULNESS

GRADE STATUS	RELATIVE HELPFULNESS		
	First	Second	Third
Accelerated* . . .	10	5	
Normal	4	11	3
Retarded	4	2	12

* There were only 15 instead of 18 possible positions for the accelerated, as there were no pupils who could be definitely classified as accelerated in the Z grade 5 group, which furnishes 3 of the 18 sets of records utilized. Hence 3 records for the "accelerated" row are missing.

Two-thirds of the accelerated, normal, and retarded groups are found respectively the most helpful, average, the least helpful, with the accelerated not once taking third place. If the fact of being too old for one's classmates operates by making one feel out of place and less inclined to do what the others do — which in itself might be one factor in causing such pupils to fall back — then it must also be said that being younger apparently carries with it a greater activity in other directions than regular class work, for the accelerated are also usually the most coöperative. We have also seen that girls are more affected than boys by age-group status. These facts, obviously, cannot be wholly understood without refer-

ence to the effect of intelligence on the tendency to be of service, to which we shall now turn.

INTELLIGENCE AND SERVICE

This is not the place to enter into a discussion of the meaning of the term "intelligence." In mental testing the word is used variously to refer to (1) intelligence test scores, (2) mental age, based on the average test scores of children of each chronological age, (3) intelligence quotient, or the ratio of mental age to chronological age, and (4) the deviation of a test score from the average of the age group to which the subject belongs divided by the standard deviation of that age group.

In *Studies in Deceit* we correlated our facts with all four types of intelligence rating and found correlations between intelligence and honesty tests ranging all the way from zero to .50, according to population and test. We figured that the probable correlation between honesty as measured by a large number of tests and intelligence in the fourth sense described in the preceding paragraph, which is roughly equivalent to the IQ, was about .60. In the case of our tests of coöperation and charity, we used only this fourth type of intelligence score. The *r*'s* between intelligence as thus measured† and the total service scores for our three populations and the total are for X, .095; for Y, .193; for Z, .163; and for XYZ, .161. By and large, intelligence has only a barely perceptible relation to the tendencies to be of service which we have measured.

Further study of the scattergrams brings out some interesting facts. The highest ten per cent on the service tests, for example,

---

* These *r*'s and some others are of only slight value, as the regressions involved are frequently nonlinear. When this is the case, we have usually run means or medians of the arrays and presented the results graphically rather than in the form of etas.

† Appendix D describes the method used in computing the sigma values of the CAVI achievement scores which were gathered for our entire population. On our scale of intelligence scores, 50 represents average for age.

have a median intelligence above 50 (which is just average), whereas the lowest eight per cent have a median intelligence rating of only 41. As we shall bring out later, our most intelligent population, X, is not the most coöperative. That honor is reserved for Y, which is far lower on the intelligence scale though still significantly above Z. With the exception of three of our twenty-two classrooms, there is, further, no relation between the mean intelligence scores and the mean service scores of the groups.

Another way of analyzing the data is to divide the children into a middle or average group of intelligence, containing about two-thirds of the cases, and two other groups, one the bright children who are above the mid-group, and the other the dull children who are below the mid-group. On our scale of intelligence, the general average of which is 50, the mid-group range is 40 to 57.9.

Figure 21 shows the extent to which intelligence is associated with service. In Y the more intelligent are more coöperative;

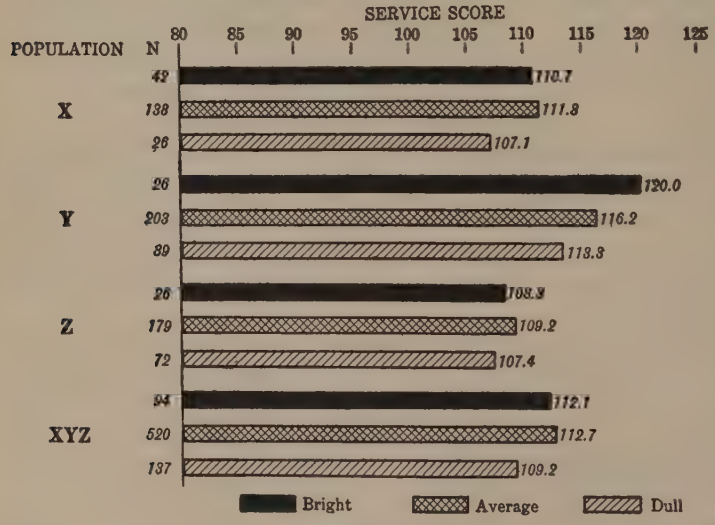


FIGURE 21

Comparison of Bright, Average, and Dull Pupils with Regard to Median Service Scores — Populations X, Y, and Z

in X and Z the average group is superior. But in all populations the dull children are definitely less helpful and charitable. This figure should be studied in connection with Figure 22, which turns the facts the other way around. The cases in the three service groups, upper, middle, and lower, do not exactly correspond in number to those of Figure 21, but the tendency for those scoring low in service to be also low in intelligence is quite clear. At the

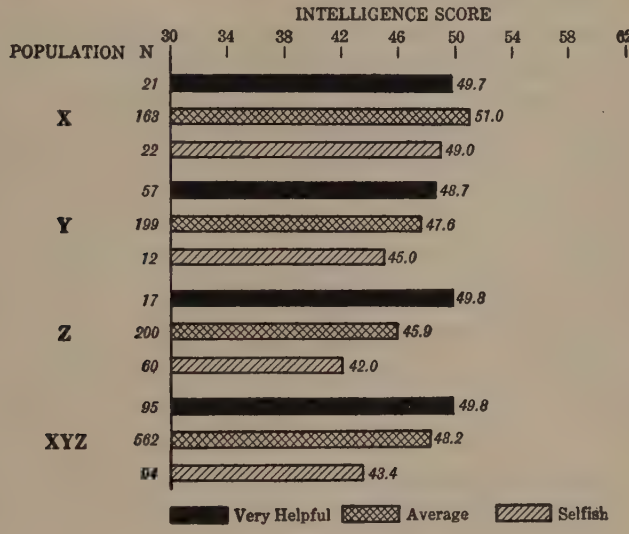


FIGURE 22

COMPARISON OF VERY HELPFUL, AVERAGE, AND SELFISH PUPILS WITH REGARD TO MEDIAN INTELLIGENCE SCORES — POPULATIONS X, Y, AND Z

upper end, however, we find the 17 Z pupils who may be called definitely bright, to be also very coöperative. So also in Y, but not in X. All in all, the most coöperative children are slightly brighter than those who show a more moderate self-sacrificing tendency.

The reason for the apparent discrepancy between Figures 21 and 22 is that, although the most coöperative are also the brightest, there are nevertheless many of the brightest children who did not do well on the coöperation tests. Since brightness obviously

does not depend on coöperation, this would mean that one condition of a high degree of coöperation is a high degree of intelligence.

Another way of presenting the facts is to say that the tendency to serve, which is low on the average among the less intelligent, increases as intelligence increases up to about average intelligence, after which an increase in intelligence does not connote a corresponding increase in service tendencies — in fact, from this point up it tends sometimes to decrease.*

When the service tests are treated separately, the relation between intelligence and the various types of helpful coöperative behavior corresponds to what we have just noted for the total service scores. The details are printed in Tables XVII and XVIII of Appendix C. The chief exceptions, on all five tests, to the tendency for the duller children to prove less helpful are: (1) the Kits test: in our underprivileged community, Z, the brighter children tended more than the others to keep the pencils, erasers, and so on, which had been given to them, rather than let them go to others; (2) the Money Vote: again in Z we find the dullest least inclined to vote the money for the more selfish purposes; (3) in Y also the dullest vote most unselfishly; and (4) on the Envelopes test the dull group in X does better than the bright group, though not so well as the average.

As there are three schools and five tests, there are fifteen records. The relative helpfulness of the bright, dull, and average groups in these fifteen instances is shown in the accompanying diagram.

NUMBER OF CASES IN WHICH THE BRIGHT, AVERAGE, AND DULL CHILDREN WERE FIRST, SECOND, OR THIRD IN HELPFULNESS

RELATIVE INTELLIGENCE	RELATIVE HELPFULNESS		
	First	Second	Third
Bright	9	3	3
Average	4	9	2
Dull	2	3	10

* As stated before, this means that the regression is nonlinear.

In conclusion, then, we find a slight tendency for the more intelligent to be more helpful and coöperative. While the differences are not great, we can say that the brightest are also very coöperative in three-fifths of the separate records we have on three populations, whereas the dullest show little tendency to be of assistance to others in two-thirds of their records; and in three-fifths of their records those whose intelligence is about average are also about average in service.

SCHOLASTIC MARKS AND SERVICE

Certain of the factors we have been dealing with may be thought of as represented in teachers' grades. Intelligence and age as factors in retardation or acceleration are necessarily reflected in the kind of work that is being done. Those who get the best marks, other things being equal, are both more intelligent and better placed in their grades. As our findings show a relationship between both intelligence and proper or accelerated grade-status, it is not surprising to discover that the correlations between teachers' marks and the total service scores are for populations X, Y, and Z, .266, .239, and .346, respectively, with .316 for the total. This indicates that, on the whole, the children who do good class work tend also to engage more fully in helpful acts as occasion arises.

It is of interest to compare with these *r*'s between scholarship and service the correlation between deportment and service. The *r*'s run .154 for X, .209 for Y, .233 for Z, and .194 for the total. It is hardly worth while to try to disentangle intelligence, scholarship, and deportment when the relation between each and the tendency to serve is so small. They are doubtless all involved, although the more directly responsible factors may probably be best understood in terms of general social adjustment. The coöperative child tends to be also the " good " child in the ordinary sense, but it is clear that conventional goodness and responsiveness to the social needs embodied in our tests are by no means identical.

The marks were taken from the school records and cover the period of the testing. As the correlation between marks and

intelligence is not very high, there is room in the association between marks and service for such factors as industriousness to enter in. In Volume III we shall take up the problems of the interrelations of the two tendencies dealt with in Book Two of this volume with the two types of social behavior we have been studying, namely, deception and service.

CONCLUSIONS CONCERNING AGE, GRADE, AND INTELLIGENCE

It has been brought out that age, grade, and intelligence are not consistently associated with tendencies to serve. Populations differ in the way in which the several tests correlate with these various facts. Generalizations therefore tend to be misleading. Apparently there are environmental factors which engulf any slight association there may be between age and intelligence and service tendencies. The duller pupils are, on the whole, less coöperative, but this may be due in general not so much to their lack of intelligence as to their retardation in school. They are not properly adjusted to their environment. It is their *relative* unintelligence, perhaps, which puts them at a disadvantage, as may be seen from our favored population, X, the mean intelligence of which is 50.8, as compared with 45.4 for Z. Yet even in X the median service score of the dullest pupils is only 107, as compared with 111 for the brighter and average groups. The age differences may be looked at in the same way. The older pupils, grade for grade, are duller and retarded.

Brightness, however, does not invariably imply superiority in service tendencies, although those who are most coöperative are usually also the most intelligent. Brightness is sometimes accompanied by a *decrease* in service scores — as though some accelerated groups also felt a little out of place. This is not the usual tendency, but the facts indicate that the child who is normal for his grade is at no disadvantage in learning to act helpfully and coöperatively.

On the whole, children do not become more helpful merely by growing older. They may become either less helpful or more according as the experience gained in their particular environment discourages or encourages coöperation and charity. The consideration of these environmental factors and their method of operation will be taken up in Chapters IX to XIII.*

* For convenience of reference the correlations between total service score and the factors studied in this chapter are quoted here, giving the separate r's for populations X, Y, and Z in this order, with a fourth for the three populations combined. For age, the r's are $-.203$, .052, $-.158$, and $-.18$; for intelligence, .095, .193, .163, and .161; for scholastic marks, .266, .239, .346, and .316; for deportment, .154, .209, .233, and .194.

CHAPTER VII

SEX DIFFERENCES IN COÖPERATIVE TENDENCIES

It is a well-known fact that consistent and significant sex differences in social and mental traits are hard to find. Boys and girls score on the average about the same in intelligence tests, school achievement tests, and practically all tests requiring mental skills. So much alike are the sexes in mental qualities that the separate education of boys and girls on the ground of differences in mental ability cannot be justified. Whatever reasons there are for separation are based on social considerations. It remains for science to determine whether there are differences between the sexes in character qualities and, if so, in what respects and how much they differ and whether these differences are important for character education.

In our previous studies of honesty we found no consistent sex differences except in the case of school tests done at home, where girls cheated more than boys. This we attributed to the fact that girls have more ambition, not less honor, than boys. In the present studies in service we find fairly consistent differences in favor of the girls, and in certain of the tests these differences are large.

SEX DIFFERENCES IN TEST SCORES

The Free Choice Test. The facts concerning this and other tests are printed in the tables of Appendix C, which may be consulted in support of the statements made in the text. As in previous chapters, we shall present the essential relations in graphic form wherever possible, reserving for footnotes any detailed or technical treatment.

In the Free Choice test, the score is the number of choices made to let one's work count for the class *instead of* for self. The difference in the *amount* given to the class varies only slightly between the sexes and is not consistently in favor of one or the other. A far larger per cent* of girls than of boys, however, gave at least one unit of work to the class. This is clearly brought out in Figure 23.

There is also a real difference between the sexes in variability. Twice as many boys as girls gave *everything* to the class and half as

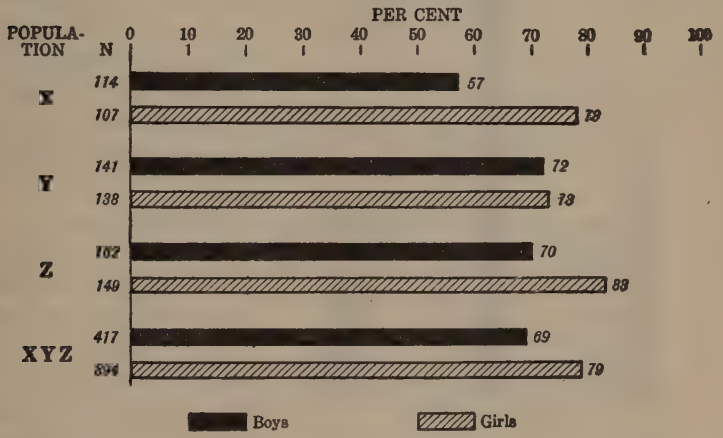


FIGURE 23

PER CENT OF EACH SEX COÖPERATING ON THE FREE CHOICE TEST — POPULATIONS X, Y, AND Z

many again, as may be inferred from Figure 23, gave *nothing* to the class. That is, the boys tended to bunch at each end of the distribution, as shown in the adjoining chart. In Figure 24 the black columns, representing the boys, are higher at the sides and lower in the middle than the shaded columns, representing the girls. More boys than girls appear both extremely unselfish and extremely selfish. Figure 25 gives the same facts for the whole scale. Had the scale been longer at each end — that is, if there

* The difference of 10% is 3.2 times its standard error (SE).

had been more opportunities to choose between self and class — probably the boys' curve would have tapered off more slowly. As in many other types of behavior, the girls are more concentrated in the middle and the boys scattered more widely up and down the scale: a far larger per cent of the girls *shared* (giving neither all nor nothing) than the boys.* On several of our tests, however, the girls are more variable than the boys.

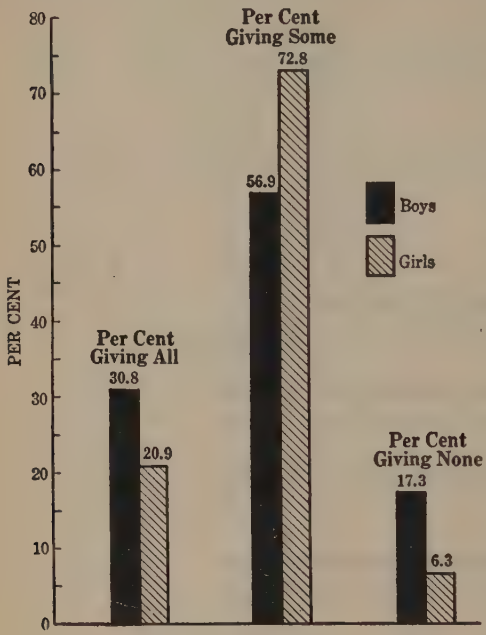


FIGURE 24

Sex Differences on the Free Choice Test

The Efficiency Coöperation Test. This test, it will be recalled, involves also the contrast between self and class; only here the demand made upon the unselfish or loyal person is one which requires endurance as well as a generous or loyal impulse. Here the boys consistently excel the girls, though only slightly,† and are on the whole *less* rather than more variable.

As this test involves six units of work for self and six for class for successive two-minute intervals, a more detailed com-

* The fact of sex difference in variability may also be seen by comparing the SD's of the two groups. Table XIX in Appendix C shows that the difference between SD's is 3.5 times its own SE, which indicates a truly significant difference in variability on this test.

† The mean difference is twice its SE but is maintained through the three populations.

parison of sex differences is possible. The facts are portrayed in Figure 26. It may be seen that during the first minutes of work the difference between work for class and work for self is less in the case of the girls than in the case of the boys. If the test had

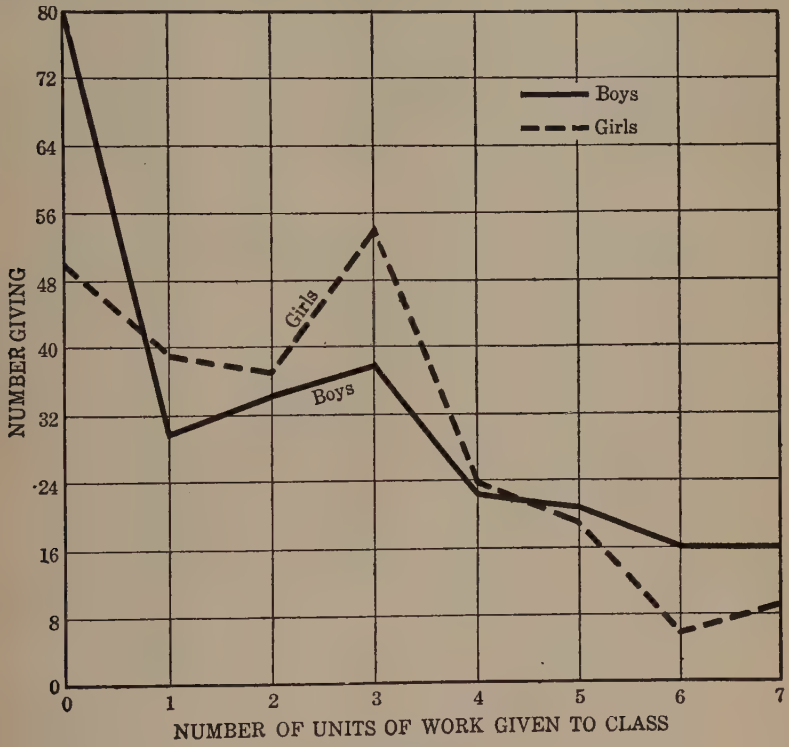


FIGURE 25

Distribution of Free Choice Scores by Sex

lasted only four minutes instead of twenty-four, the girls would have appeared more coöperative than the boys. But as the test proceeded, the girls worked relatively less and less hard for class than for self, until in the last four minutes the difference in favor of the boys is considerably greater than it is on the average.

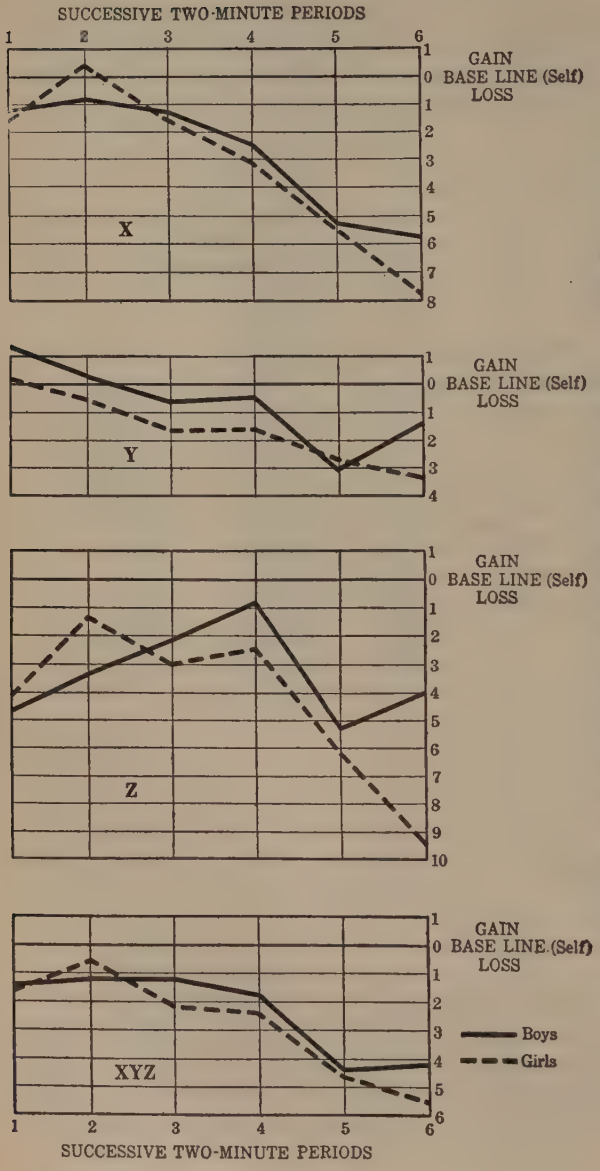


FIGURE 26

SEX DIFFERENCES DURING THE EFFICIENCY COÖPERATION TEST

The number of examples done for self is represented by the straight line. The extent to which the work was done for class during each successive two-minute period is shown by the curve. As the test progresses the number of examples done for class decreases more and more rapidly in comparison with the number done for self.*

* The figures for this chart are in Table XXI in Appendix C.

This difference may be further analyzed to bring out the fact that the superiority of the boys is even greater than the scores show. Figure 27 reveals that the work done for self by the girls remains about the same throughout the twenty-four minutes, but the amount they do for class falls off rapidly, thus showing a constantly increasing difference. The larger the gap between the two curves on the diagram, the larger are the scores for self. In the case of the boys the self scores do not hold up as well as those of the girls, but the class scores *do not decrease so fast as those of the girls* and the

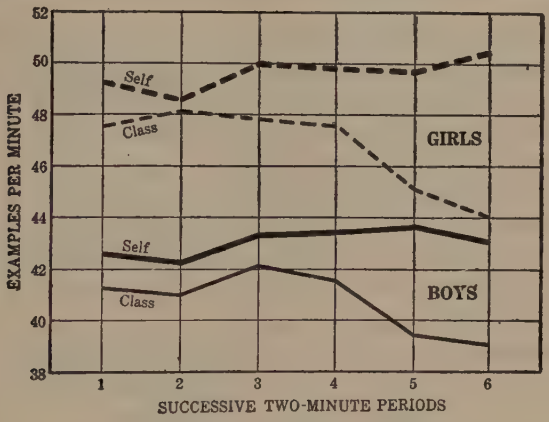


FIGURE 27

SEX DIFFERENCES IN THE AVERAGE NUMBER OF EXAMPLES PER MINUTE DONE FOR SELF AND FOR CLASS IN EACH TWO-MINUTE PERIOD OF THE EFFICIENCY COÖPERATION TEST*

difference between self and class is therefore less. If the curve of the boys for class did not hold up as well as it does, we should have to regard the superiority of the boys as a fallacy resulting from the nature of the score; but even if the boys had held out for themselves as well as the girls did, they still would have shown relatively superior coöperation scores. They worked relatively harder for class than the girls, but not so hard, relatively, for themselves.

* See Table XXII in Appendix C for supporting data.

The Money Vote Test. The scores on the Money Vote test are so adjusted as to rate a vote to give some prospective class prize money to an outside charity higher than a vote to give the money to school or class. Scored in this way, the girls appear consistently more generous than the boys. There is about one chance in a thousand that, if this test were repeated, the boys would exceed the girls.

The Kits Test. In per cent of cases giving away at least one of the ten objects contained in the school kits, the boys and girls are about alike.

In this respect the Kits test affords an interesting contrast to the Free Choice test. In the latter, where the *class* was the occasion for self-denial, more girls than boys proved coöperative. In the former, where children in other places were the occasion of giving, the boys and girls gave in about equal numbers.

In the *amount* given away, the situation is about the same as for the Free Choice test, where there was an opportunity to give away one's chance to work for himself. As in that test, the boy and girl givers gave about the same amount, but the boys were consistently, though only slightly, more *variable.*

The Envelopes Test. This test requires the collection of pictures, stories, and jokes at home, and as might be expected the girls come forward in larger numbers to give some help. Of course this does not *necessarily* imply any fundamental difference in responsiveness to the needs of hospital children, but it might be so interpreted were it not for the fact that the Kits test, which provides the material to be given away and requires immediate decision, shows no sex differences either in numbers giving or amount given. The sex differences on the Envelopes test, which are quite beyond the limits of chance,* are shown in Figure 28. In the *amount done* by those who did anything, the girls again exceed the boys except in one population — the favored city community known as X. Here the two sexes are alike. In variability, the girls exceed the boys on this Envelopes test.

---

* For populations X, Y, and Z combined the difference is 5.7 times its SE.

Summary. These detailed differences in the charitable and coöperative tendencies of the two sexes may well be summed up before we present the case for the total service scores. On the two tests which require immediate choice and action in the classroom, namely, the Kits and Free Choice, the girls and boys are indistinguishable in the amount of helpfulness shown, but the boys vary more among themselves than the girls do. Also more girls than boys respond to the appeal to help out the class. On the Money Vote test, where there is a choice but no immediate action,

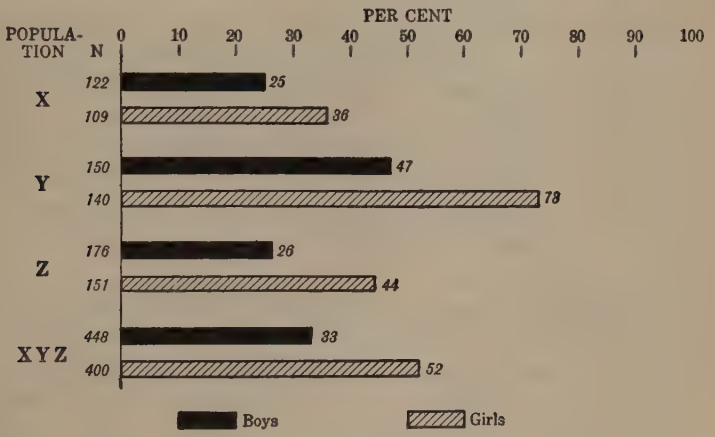


FIGURE 28

PER CENT OF EACH SEX COÖPERATING ON THE ENVELOPES TEST

the girls show a superior tendency toward charitable decisions regarding the way in which a sum of money should be spent. On the Envelopes test, which was done at home, the girls both are more commonly helpful and also do more work than the boys. On the Efficiency Coöperation test, the girls at first exceed the boys, but they soon drop off in the rate at which they work for their class, whereas the boys, although less persistent for themselves, keep their work for class at a rate relatively superior to that of the girls.

The Total Service Score. As might be anticipated from our discussion of sex differences on the separate tests, when we combine

the scores into a total service score, the boys and girls turn out to be much alike, with the girls slightly more coöperative* in each population. The difference for the whole population is 1.7 points on a scale which has a mean of 112 for all cases and which ranges from 60 to 165. The chances are 97 in 100 that the girls would always score higher than the boys.

The girls are more variable, except for population X, and the difference of 1 point in variability (SD) between boys and girls for X, Y, and Z together is well within the limits of chance.

SEX DIFFERENCES IN REPUTATION SCORES

Although this chapter deals primarily with test scores, the sex differences noted show such interesting parallels with those found in the reputation scores as to warrant our giving space to some of the more outstanding facts. The reputation score for populations X, Y, and Z was composed of three sets of data, constituting those aspects of the children's reputation which we could reduce to quantitative terms for the entire population.

Sex Differences on the "Guess Who" Test. Large and consistent differences appear in the way the members of a single group vote for boys and girls respectively. On the ten items of the "Guess Who,"† dealing with such things as lending a hand, sharing one's property, helping out the group, the girls secure far higher ratings than the boys.‡ Figure 29, with its associated table, No. XXVII in Appendix C, shows that the per cent of girls receiving plus votes is 85 as against 57 for the boys. The table in the Appendix also indicates that the average score for the girls is + 4 and for the boys about zero — in population Z it is actually −.64.

The boys vary among themselves also considerably more than do the girls, which would indicate that there was in the group as a whole a greater difference of opinion about the boys than about the girls. A clue to this situation was found by consulting the

* The difference of 1.7 points is 1.9 times its SE.

† See Chapter IV for details.

‡ The difference in means is 7.9 times its SE for all cases combined.

records of two classes for which the names of those voting were known. The girls in these two groups rated the boys lower than they rated girls, whereas the boys rated the girls on the whole higher than they rated boys.

The Check List. The teachers checked more girls than boys with words indicating positively coöperative tendencies, as shown in Figure 29. On this device each pupil, it will be recalled, received a plus, minus, or zero score representing what two teachers said about

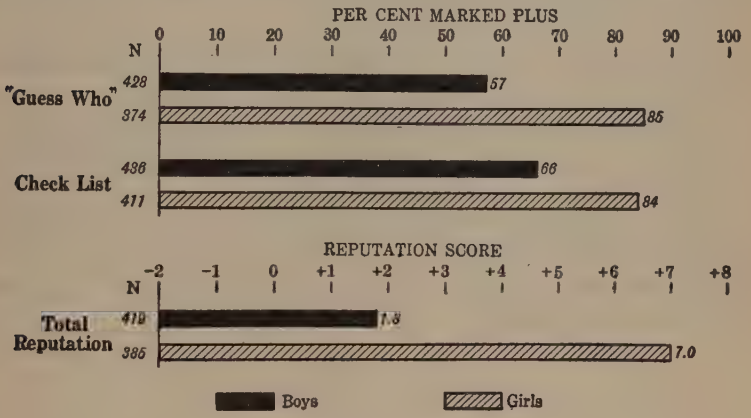


FIGURE 29

SEX DIFFERENCES IN REPUTATION SCORES — POPULATIONS X, Y, AND Z COMBINED

him on two separate occasions apiece. The percentages given are for those receiving a plus score.

These differences are about as extreme * as in the case of scores on the " Guess Who " and are consistent for the three schools.

The Total Reputation for Service. As similar differences would be found on the other technique, we may turn at once to the total reputation score, which shows even more conspicuous sex differences. The means and per cents show similar results, but we have given the means in Figure 29 for the sake of comparison with

* 6.2 times the SE for all cases combined.

previous figures. From Table XXVIII in Appendix C it may also be seen that the boys are consistently and extremely more variable in reputation than the girls.

Girls and women have always been regarded as more likely to prove self-denying and helpful than boys and men in situations appealing to pity or sympathy, and this tradition is reflected in the reputation scores we have reported. The difference between the sexes in reputation, however, is so much greater* than it is found to be on the tests that one is inclined to believe that a sex prejudice may be at work which rates the boys lower than they really are and the girls higher than they really are. By tradition teachers are partial to boys, but the evidence of our results is in favor of an opposite view,† so far as this type of conduct is concerned.

Further evidence of the existence of some partiality was brought forward in the footnote on page 104. Here it was noted that the r's between the reputation scores and test scores for boys ran 60% higher than the corresponding r's for girls in populations X and Y. This is what would happen if an element of prejudice were inserted into the reputation scores of the girls.

GRADE AND SEX

Before reaching our final conclusion as to sex differences, we should at least raise the question of the relation of age to such contrasts between boys and girls as we have reported. If the biological differences between boys and girls of the ages with which we are dealing, namely, 9 to 15, are complicating the situation, this ought to appear if we report the scores by both grade and sex for each population. From Table XXIX of Appendix C it may be seen that the total service scores are with two exceptions higher

* The difference for the total service score is 1.9 σ (*i.e.*, 1.9 times its SE) and for the total reputation score 7.9 σ, or more than four times as much.

† In population X, where sex differences are slight, the difference between the sexes in reputation corresponds exactly to the difference in actual scores. Indeed the difference in each case is 1.54 σ, or 1.54 times the SE of the difference.

for girls in each school for each grade. The differences between boys and girls, however, are larger for certain grades than for others and vary among the schools, which indicates the need for a more detailed age-sex analysis than our time has permitted.

CONCLUSIONS

It is apparent that in most tests, and on the whole, the girls are more helpful than the boys and are so regarded by their teachers and girl friends. In some cases we find more boys than girls who are extremely generous or extremely selfish. The differences, however, are not so uniform or so great as to justify a sweeping generalization. They may be due to the particular types of situation included in the tests. As we have already pointed out, coöperative behavior is specific and largely a function of the situations in which children are placed and in which they have acquired certain habits or tendencies. The sex differences reported, therefore, should be thought of in connection with situations which are similar to those in which such differences have been observed.

CHAPTER VIII

PHYSICAL AND EMOTIONAL CONDITION IN RELATION TO THE TENDENCY TO BE OF SERVICE

PHYSICAL CONDITION

It is commonly believed that a child's conduct is conditioned in no small degree by his health. That defective eyesight or hearing is a handicap in school there can be no doubt. Other physical defects, such as cardiac trouble, have their influence on conduct and social attitudes. The relation of general health and all-round physical condition to conduct is also of great interest to all who are attempting to control the process of character formation. A great deal of scientific work has already been done in this field, which we shall not attempt to summarize since we have not endeavored to deal with this problem in any exhaustive manner. The data we have collected, however, may have some importance, as we find only a slight relation between physical condition and the tendency to be of service.

In *Studies in Deceit* we described how we secured measures of certain types of physical strength in connection with one of our series of honesty tests, which was administered as an athletic contest. The four measures obtained were: (1) strength of grip (one hand), three trials with a Smedley hand dynamometer; (2) lung capacity, three trials with a standard wet spirometer; (3) standing broad jump on a standard mat; (4) number of pull-ups (chinning) for boys and number of times the knees could be raised while hanging suspended from the bar for girls. (The knee-raising test was introduced for the girls because few of them could chin themselves even once.) The scores on these tests we combined into

a single index of physical ability and another index of condition, keeping separate the boys and the girls.*

* Table XXVIII gives the essential facts.

TABLE XXVIII

BASIC PHYSICAL ABILITY SCORES

	Mean	SD	Weight	Weighted SD
Boys				
Grip	20.3 kilograms	8.32	1	8.32
Lung Capacity	139.2 cubic inches	35.9	$\frac{1}{5}$	7.19
Pull-ups	172.0 times $\frac{1}{10}$ weight	30.0	$\frac{1}{3}$	10.00
Broad Jump	62.5 inches	7.68	1	7.68
Girls				
Grip	17.7 kilograms	7.70	1	7.70
Lung Capacity	128.3 cubic inches	27.2	$\frac{1}{5}$	5.44
Knee Raising	18.7 times	10.0	$\frac{1}{2}$	5.00
Broad Jump	55.5 inches	7.34	1	7.34

First it may be noted that the tests are not equally weighted according to the SD's. In the case of the girls the relatively low weight given to Lung Capacity and Knee Raising is due to the fact that the tests are not regarded to be as valid as the others. Rogers ("Physical Capacity Tests in the Administration of Physical Education," *Teachers College Contributions to Education*, No. 173, 1925) showed that lung capacity has a low correlation with athletic index, and he gave it a low weighting. We have deliberately weighted it lower for both boys and girls. In the case of the boys, chinning is weighted highest because we regard it as the best test. Here we multiply the number of pull-ups by $\frac{1}{10}$ the weight. In the present study we used Rogers' formula of

Chinning score = Chinning times $\frac{1}{10}$ weight − Average height of group.

But we omitted the height fact altogether partly because Rogers shows that almost as valuable results may be obtained by omitting it and partly because we did not find height an important factor.

About 25% of the boys did not chin themselves once. Hence their chinning scores are zero. In these cases we predicted the chinning score from their jump record, as we did in *Studies in Deceit*, except that we use a new formula for our present data, which is: Chinning = .36 (jump in inches) − 20, almost the same as the former one, which was Chinning = .30 (jump in inches) − 16.

Using the weights in Table XXVIII, we summed the highest practice scores on the four tests. This score was taken as the physical ability score.

Rogers has shown that physical ability as determined here depends on age and weight. For this reason the scores obtained by children of different ages and weights are not comparable, just as point scores on intelligence tests are not comparable for children of different ages. It is required, then, to derive an index that will indicate each child's physical fitness for his age and weight. This was done in something the same manner as in the case of our point scores on the CAVI intelligence tests. The actual scores were converted into ratings for each age group, except that here we considered weight as well as age. For example, we found the average physical ability score of all boys of 12 years of age who weighed from 100 to 110 pounds. This was taken as the norm for all such boys, and the score made by each boy of this age and weight was divided by this norm. The average boy scored 1 or, to avoid decimals, 100. In this manner a " physical condition " index was obtained for each child, showing his physical fitness for his age and weight. Such scores ranged from 40 through 100, or the average, to 160, in a way comparable to IQ's. The correlations between this index and the total service score for populations X, Y, and Z and all populations combined are, respectively, .082, .019, .134, and .018. Again these low *r*'s may be due in part to heterogeneity in the sources of the data. A complete analysis might show that for certain groups the correlations would run higher. However, these physical tests are after all strength tests and do not indicate subtle physiological variations which may be correlated with certain types of conduct.

Some of these more subtle factors were included in an investigation conducted throughout the country in 1927 to 1928 by the American Child Health Association.* We were fortunate enough to secure their coöperation in the examination of about 175 of the children in population Y. The major facts observed were nutrition, eyesight, condition of teeth, and cleanliness. The teeth score consisted of the number of uncorrected defects. The nutri-

* For detailed description of the techniques used those interested are referred to the American Child Health Association, 370 Seventh Avenue, New York City.

tion score was determined by weight and the amount of superficial tissue over certain muscles. The eye score was a more conventional measure of vision. A highly standardized technique was developed for the conduct of the examinations, resulting in scores of high reliability and objectivity. The results give us data which nicely supplement what we ourselves secured in the athletic contest just referred to. Our tests measured condition as expressed in or indicated by anatomical status, whereas these other tests measured condition as expressed in or indicated by physiological status.

In order to get some sort of combined test from the material provided us by the Child Health Association, we weighted the scores on teeth, eyes, and nutrition by their respective SD's and added them to make a physical condition total. This we then correlated with our total service score and each separate test. The resulting r's, all low, are as follows: with total service score, .17; Free Choice test, .01; Efficiency Coöperation, − .15; Money Vote, .13; Kits, .26; Envelopes, .17. It is worth noting, however, that all the children of the upper 5% in the total service score are practically without physical defects in eyesight, nutrition, or teeth.

The separate tests in the physical condition battery were also correlated with the total service score, and we found that the correlation between vision and service was − .05; between nutrition and service, .10; and between condition of teeth and service, .12. Had the eye test been omitted from the total, the correlations between service and physical condition would have been somewhat higher. The median service scores for three levels of healthy condition of teeth and nutrition are shown in Figure 30.

Conclusions. Evidently physical condition as measured by the two types of test used has only a slight relation to the tendency to be of service. For the test based on strength, the r's with service are .082, .019, .134, and .018 for X, Y, Z, and all combined. For the test based on status, the r in Y is .17. At the upper extremes of condition and of helpfulness, however, we do find a strong tendency for fine physical condition to be associated with a high degree of unselfishness. Considering teeth and nutrition by themselves,

the relation, though slight, is positive in each case. Such matters probably are associated also with cultural level and intelligence, so that after allowing for their influence little evidence would be

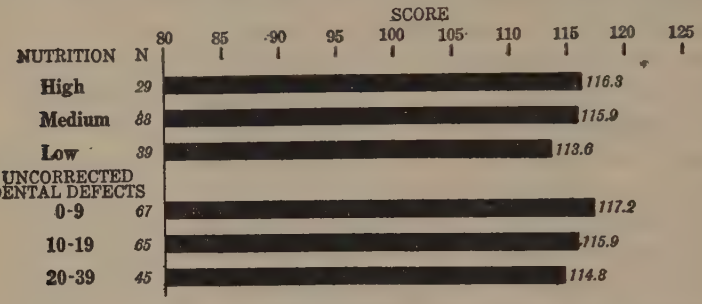


FIGURE 30

CONDITION OF NUTRITION AND OF TEETH IN RELATION TO MEDIAN SERVICE SCORE

left for the presence of a physiological factor in the types of behavior we call charity and coöperation.

EMOTIONAL CONDITION

The problem of the social and emotional adjustment of the school child to his home, school, and other environment has attracted the attention of psychiatrists, sociologists, and educators. Much work has been done on the problem and valuable results have been achieved, particularly by clinical and case study methods. Only a few tests are available.

In laying out the plan for the present battery of character tests we have endeavored to tap this particular side of character, using for this purpose the Woodworth-Mathews Personal Data Sheet.* This was supplemented by a list of questions of our own which forms part of a general questionnaire called the Pupil Data Sheet.† On those cases in Population Y who were given the health tests we

* Published by C. H. Stoelting Co., Chicago.

† Printed in full in Appendix B.

have also reliable evidence concerning the practice of nail biting. Thus we have, in all, three measures of emotional condition.

Results of the Woodworth-Mathews Personal Data Sheet. The history and nature of this instrument is known to students of this subject and need not be repeated here. Reference should be made, however, to an article by Dr. Mathews, describing the questions and the scoring plan with tentative norms.*

Some general notion of the type of question asked may be had from the following samples chosen at random from her list of questions:

Do you like to play with other children?	Yes	No
Do you feel that nobody quite understands you?	Yes	No
Do you find school a hard place to get along in?	Yes	No
Are you afraid of the dark?	Yes	No
Can you sit still without fidgeting?	Yes	No

There are 75 questions such as these, and each is answered by underlining Yes or No. The score on the test is the number of "unfavorable" responses. An unfavorable response is one that indicates maladjustment or neurotic tendencies. For example, in question 1, the unfavorable response is No; in question 2 it is Yes; in 3 it is Yes; in 4, Yes; and in 5, No. The higher scores thus indicate greater emotional difficulty.

Dr. Mathews gave this test to 1133 school children, mostly in New York public schools. They were in grades 4 to 12, inclusive, and their ages ranged from 9 to 19. The results constitute her tentative norms, which she reports by age, grade, sex, nationality, and school retardation or acceleration. For the entire group tested (both sexes) the median number of unfavorable responses is 23; for the boys alone it is 20, and for the girls, 25.5. She reports also interesting nationality and age differences, for quoting which we cannot allot space here. We shall, however, compare the results of the tests on our populations with her figures wherever such com-

* Mathews, Ellen, "A Study of Emotional Stability in Children," *Journal of Delinquency*, 1923, **8**, 1–40.

parisons are pertinent. Table XXIX presents a comparison of the medians of our three populations by sexes compared with Dr. Mathews' medians.

TABLE XXIX

COMPARISON OF MEDIAN SCORES ON THE WOODWORTH-MATHEWS TEST WITH THE MATHEWS TENTATIVE NORMS

POPULATION	NUMBER OF BOYS	MEDIAN, BOYS	NUMBER OF GIRLS	MEDIAN, GIRLS	TOTAL	MEDIAN TOTAL	PER CENT AT OR ABOVE MATHEWS' NORMS	
							Boys	Girls
X	110	13	107	14	217	13.8	32%	17%
Y	138	15	132	17	270	16.2	34%	24%
Z	145	22	143	22	288	21.6	52%	43%
Norm	496	20	512	25.5	1008	23		

On the whole, our populations seem to show fewer unfavorable responses than Miss Mathews'. One reason is that populations X and Y have a greater proportion of younger children and a smaller proportion of foreign born (especially Italians) than Miss Mathews' population. She reports a median of 36 for 188 Italian children, 20 for 199 Jewish children, and 16 for 87 of mixed stock (mostly Anglo-Saxon, Celtic, and Teutonic).

After making allowances for differences in age, nationality, and school status, our results are in substantial agreement with those of Dr. Mathews. As an illustration of this statement our twelve-year-old Italian boys who are retarded one year in school show about the same number of unfavorable responses as her twelve-year-old Italian boys who are one year retarded. This is confirmation of the approximate correctness of her norms. The chances are that if other samplings of twelve-year-old Italian boys one year retarded in school were taken their median would be about as those we have reported.

The correlations between the emotional stability score on this test and our total service score for populations X, Y, Z, and all

combined are .082, .172, .068, and .152, respectively. There is only the slightest tendency for the more stable and adjusted to be more helpful. The correlation scattergram shows 23 pupils out of the entire 707 tested who scored above 135 on the service tests. Of these 23 all but one are above the average in stability. Of the 42 cases scoring below 95 on the service tests 23 are below the average in emotional stability. Thus it appears that the extremely helpful and coöperative children are reasonably sure to be emotionally stable, but that the chances are only about 50–50 that an extremely selfish child will be neurotic.

Results of the Pupil Data Questions. The second measure of emotional condition is based on a series of questions asked in our own Pupil Data Sheet. The first edition of this questionnaire appears in Appendix I of *Studies in Deceit.* In the second edition numerous changes were made. The questions that provide the basis for the emotionality score are as follows:

What are you most afraid of?

Which teachers have you disliked or not got along with?

Are you ever punished for doing something you know you ought not to have done? About how often? By whom? What kinds of punishment do you the most good? What kinds of punishment do not do you any good?

The total score consisted of the number of apparently bona fide fears named, plus the number of teachers disliked, plus the number of punishments per month. The scores run from zero to sixteen and are distributed in our three populations as shown in Figure 31. The curve shows that fifty per cent of the cases scored either zero or one. The other half of the population are distributed from 2 to 16 in a manner that resembles one-half of a normal curve. Apparently, then, this test measures only one-half of the population, the half that is the most maladjusted. The correlations with tests are therefore somewhat lower than they would be if the entire population were measured.

The obtained *r*'s for X, Y, Z, and the total are .06, .032, −.095, and −.017, respectively. When these *r*'s are corrected for the

restriction in the range of the neurotic scores which we have just pointed out, they are about doubled, but they are all so near zero that even 100% increase does not raise them enough to make it

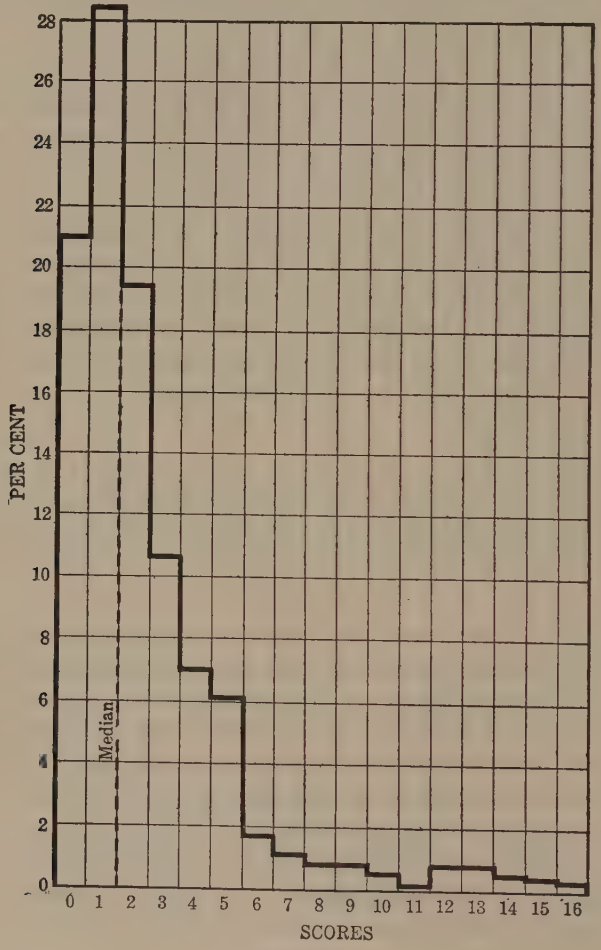


FIGURE 31

PERCENTAGE DISTRIBUTION OF NEUROTIC INDEX SCORES BASED ON PUPIL DATA SHEETS

evident that these two variables are correlated to any important degree.*

When only the extremely helpful children are considered, every one of the twenty-one whose scores are 135 or over scored 0, 1, 2, or 3 on this neurotic index, only four scoring 3. This confirms the opinion just expressed, that extreme unselfishness is usually associated with a fair degree of emotional stability.

Nail Biting and Helpfulness. Among the observations made by the team of examiners who gave the health tests of the American Child Health Association was evidence of nail biting. Our records show that of the 175 children studied about 17% were nail biters. Those who had this habit usually practiced it on all ten nails. We have thus two clean-cut groups. Figure 32 summarizes the contrasts between them in the various service scores, in relation to the average scores for the whole group of children included in the investigation. The biters are shown by the black bars and the non-biters by the shaded bars. The average (median or per cent, as the case requires) is indicated by the narrow line between them.

The correlation is low, but in four out of five tests there is a small difference in favor of the non-biters. The case becomes still clearer if we consider only extremes, as before. None of the upper 6% in service, those scoring over 140, and only one of the 13% scoring over 130 on the total service score bites his nails; only one of the upper 28% of those scoring over 50 on the Envelopes test is a nail biter.

All this is in line with our other indicators of neurotic condition and confirms our findings regarding the persistent, though slight, tendency for unselfishness and emotional stability to go together — a tendency which is particularly clear among the most unselfish children.

---

* By finding the SD of the measured half of the population on the neurotic test, the SD of a hypothetical distribution of scores representing the total measured population may be predicted. By knowing the ratio between these two SD's Kelley's formula 186 (*Statistical Method*, p. 225) will apply and the correlations between total service score and neurotic score may be predicted for a total measured population.

Conclusion. We conclude that, with the exceptions noted in the case of the extremely unselfish children, the tendency to engage in helpful and coöperative acts has only a slight relation to emotional stability as measured by our tests. It is quite possible that closer relations are concealed by our having dealt with both sexes and all ages together, but time has not permitted us to make a special study of the association of emotionality and service by sex and age.

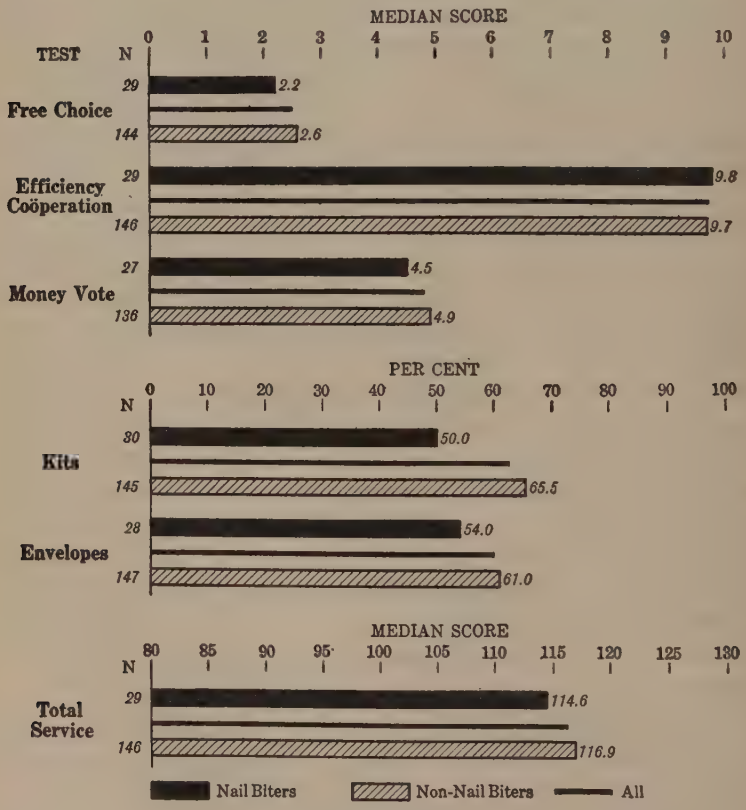


FIGURE 32

Comparison of Service Scores of Nail Biters and Non-Nail Biters

CHAPTER IX

GENERAL ENVIRONMENT AS A FACTOR IN EFFECTIVE SERVICE

COMMUNITY DIFFERENCES

In the preceding chapters we have come now and again on wide differences among the schools tested. In many instances these differences have exceeded those due to sex, grade, age, or intelligence. As we have reported all results by schools, these school differences of course do not confuse or vitiate any of the statements so far made, but they are of sufficient amount and importance in themselves to justify independent treatment.

The particular aspect of the matter with which this section is concerned is the setting of the schools in their respective communities. Some of our data are direct and some inferential. We know the occupation of the parents of the children. We have a socio-economic score, based not only on occupation but upon a large number of facts about the family life. We have also a culture score indicating the subject's ability to answer questions involving levels of cultural experience. In addition we know the religious and national background of each group and the more external features of its geographical community. Other relevant facts are such matters as intelligence of the parents, emotional stability of the children, their intelligence, their physical condition, and all the behavior differences noted on our various tests.

Such a maze of influences cannot be untangled. Nevertheless, some light may be thrown on altruistic tendencies by an examination of population differences.

Let us begin with a more detailed description of our major populations, X, Y, and Z. The children in X represent the upper

"middle class" in New Haven, Connecticut, a city of 180,000. Two schools are included. They are privileged children in the sense that they have not only the bare necessities of life but also many cultural opportunities. They live mostly in single houses with yards. The first school, grades 5 and 6, is almost like a private school in appearance and atmosphere. The children are all neighbors. The second school, of which we used grades 7 and 8, is in an older building, and there has begun an infiltration from lower economic and social levels. Both schools, however, cover a relatively narrow portion of the upper social scale and are in this sense homogeneous.

Population Y consists of all the children of grades 5 to 8 in the town of Walden, New York, which is some sixty-five miles from New York City. There are about 5000 inhabitants, mainly of English ancestry. A manufacturing plant employs a large number. All social classes are represented, but the bulk of the population is in the center of a social scale, and few families are at either the top or the bottom. There is a certain community feeling also, which welds them together in thought and standard. In this sense this group is homogeneous, but it is grouped around the middle of the scale.

Population Z is another school in New Haven, in an underprivileged and congested district. About two-thirds of the children are of foreign-born parentage. Although in some respects almost the equivalent of Y in social status, the conditions of life in Z are in extreme contrast with X and cover the lower portion of the scale.

Population P is a private school in New York of even higher socio-economic level than X.

The facts as to these differences are observational. More precise data concerning age, intelligence, and so on are printed in Tables XXX to XXXVI in Appendix C, which will be only summarized here.

Nationality and Religion. The national and religious constituency of the three populations is extremely varied. The parents in X are three-quarters native born, in Y four-fifths, and in Z one-third. This native group includes families of which one parent

was born abroad. In this mixed group, containing 21 children in X, more than half are Slavic. In Y about half are English, and in Z they are German, Italian, and Irish. Eight children, or about one-third of the small foreign group in X, are Jewish, but there are also seven children of parents born in Italy. The foreign parents of Y are chiefly German, Austrian, and Italian. In Z they constitute two-thirds of all and are 41% Italian, 28% Slavic, and 15% Irish, with the rest scattering.

As for religious affiliation, we have records for all the parents of X, for about two-thirds of Y, and three-fourths of Z. As would be expected, the Protestant group predominates in X and Y, having 49% and 61% of the children respectively. The Catholic group constitutes, on the other hand, 30% of X, 18% of Y, and 78% of Z, which has almost no Jews and only 15% Protestants. The Jewish children are found chiefly in X, furnishing 21% of the total. There are only six Jews among the children of population Y.

Age. Grade for grade the X children are the youngest and most homogeneous. The Y and Z groups are about the same age, but the Z's include a narrower range. (Y, it will be recalled, includes *all* the children of these grades in the community, but in Z we used only one fifth-grade class and one sixth-grade class.)

Intelligence. Our intelligence scores are analogous to IQ's in meaning but not in size. They mean "intelligence for age." Normal is 50. The X children are the most intelligent and, as might be expected, also the most variable. Y is next in intelligence and Z last. Y is also the least variable. Each of these mean differences is very great — that is, the three populations are really very different in the general intellectual ability measured by the CAVI tests.*

Emotional Stability. The devices used for measuring emotional condition are described in Chapter VIII. One was the Woodworth-Mathews Personal Data Sheet and the other a series of questions on our own Pupil Data Sheet. On each of these tests X stands

* The CAVI tests used are those described in *Studies in Deceit*, Book One, pages 52–53. See also Thorndike, E. L., *The Measurement of Intelligence*, 1927.

out as distinctly more stable than either Y or Z, and on the first test there is also a great gap between Y and Z in favor of Y.

Physical Condition. As described in Chapter VIII, physical ability and condition were measured by four tests of our own. In population Y several tests were administered by a team of examiners representing the American Child Health Association.

Our tests consist of measures of strength of grip, lung capacity, broad jump, and pull-up (or knee-raise). In strength, wide population differences appear even after due allowance is made for differences in age. Z, surprisingly, is at the top, and X is superior to Y in spite of the fact that the children are younger.

This low score in Y is shown even more conspicuously in the physical condition score, which is comparable to an IQ in intelligence measures. This score is physical strength for *age and weight.* The order of schools remains Z, X, Y.

Socio-economic-cultural Level. The techniques used to measure these factors are described in *Studies in Deceit*, Book One, Chapter VIII. In addition to occupational differences we have the Burdick Apperception Test, as a measure of cultural background, and the Sims Score Card, as a measure of socio-economic level. The latter could be used, however, only in X and Y. X is far above the others on all three counts, with Y somewhat higher than Z. The distributions shown in Figures 33 and 34 will make the amount * of these differences clear. Statistically, the difference between X and Y on the Burdick test is eight times as great as the difference between Y and Z. The Sims score also shows as great a difference† between X and Y.

Occupation is well known to be a good index of socio-economic and cultural level. It represents, in a way, the limits of opportunity. We classified the occupations of the fathers in five levels, following the Sims plan as described in his manual.‡ The per-

---

* On the Burdick test, X − Y is 15 times its SE and Y − Z is 1.8 times its SE.

† On the Sims, X − Y is 15 times its SE. From data gathered by Dr. Sims, we judge the difference between Y and Z to be not less than 3 times its SE.

‡ Sims, V. M., *The Measurement of Socio-economic Status*, Public School Publishing Company, 1928.

centages in each population falling at each level are shown in Figure 35. Again Y stands between X and Z but is nearer Z than X.

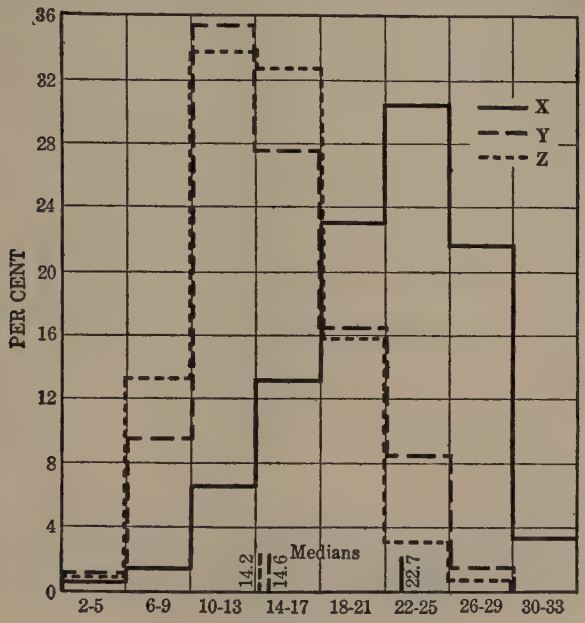


FIGURE 33

PERCENTAGE DISTRIBUTION* OF BURDICK CULTURE SCORES — POPULATIONS X, Y, AND Z

* The percentage distributions from which Figure 33 is drawn are as follows:

Scores	Populations X	Y	Z
2–5	0.5%	1.2%	1.0%
6–9	1.5	9.5	13.3
10–13	6.6	35.4	33.7
14–17	13.1	27.5	32.7
18–21	23.0	16.4	15.9
22–25	30.4	8.4	3.0
26–29	21.6	1.5	0.7
30–33	3.3		
Medians	22.7	14.6	14.2

Y covers a good deal of territory and is a complete community, and, as one would expect, there is a wide range of social level, reaching, however, neither so high as X nor so low as Z. X, on the other hand, is a section (really two sections) of a larger com-

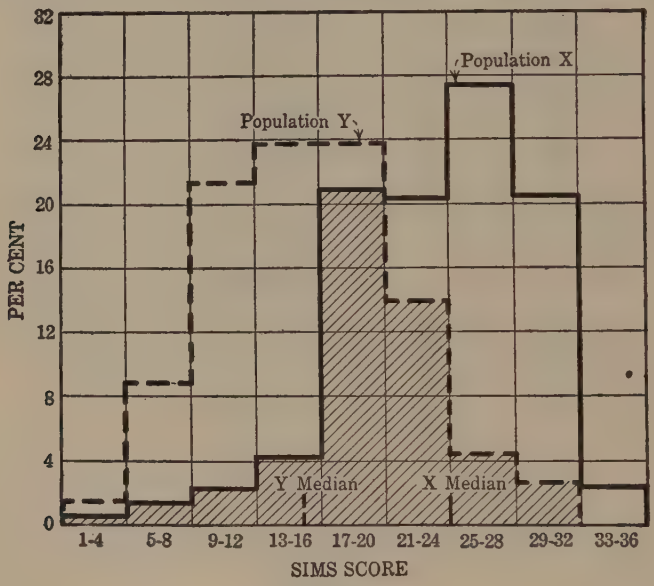


FIGURE 34

PERCENTAGE DISTRIBUTION* OF SIMS SCORE FOR POPULATIONS X AND Y ONLY

* The percentage distributions from which Figure 34 is drawn are as follows:

Scores	Populations X	Y
1–4	.5%	1.5%
5–8	1.4	8.8
9–12	2.3	21.3
13–16	4.2	23.8
17–20	20.9	23.8
21–24	20.4	13.9
25–28	27.4	4.4
29–32	20.5	2.6
33–36	2.3	
Medians	25	16.1

munity and is confined more nearly to the upper portion of the scale. Z is not confined as much to the lower part as one might expect.

The relations can best be summarized by figuring the amount of overlap of one population on the other. If we assume that Z's scores on the Sims Score Sheet follow about the same distribution as in the case of occupation, then the extent to which X exceeds Y and Y exceeds Z on all three measures — Sims, Burdick, and occupation — can be averaged. We figure that, on the whole, only 10% of population Y and only 5% of Z reach or exceed the average cultural level of X, whereas 30% of Z reach or exceed the average cultural level of Y.

In community consciousness, we might expect greater unity in Y than in either X or Z, which are in a city environment and develop more individualistic family experiences. The pupils of X and Y are also probably less associated with one another outside the schoolroom.

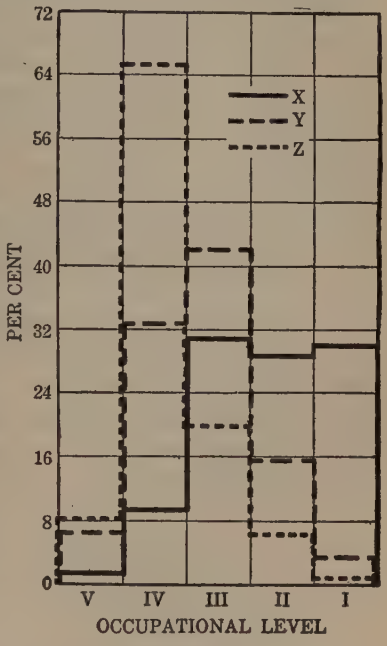


FIGURE 35

PERCENTAGE DISTRIBUTION * OF OCCUPATIONAL LEVELS — POPULATIONS X, Y, AND Z

* The percentage distributions from which Figure 35 is drawn are as follows:

Level	Populations X	Y	Z
Low V	1.4%	6.5%	8.1%
IV	9.2	32.5	65.2
III	30.9	42.0	19.8
II	28.5	15.6	6.3
High I	30.0	3.4	.7
Medians in level	2.7	3.7	4.4

School Status. *Scholastic Average.* Although the grades used are those reported by each school, the three populations nevertheless maintain the usual difference, with X at the top and Z at the bottom. The widest differences among the pupils are found in Z.

Deportment. The marks for good conduct place X at the top, with Y and Z just alike. Both observation and hearsay would place Z far lower than Y, but these differences are not reflected in the marks. The standards of conduct are obviously not comparable.

Reputation for Service. (Tables XXXVII and XXXVIII, Appendix C.) The Y pupils think somewhat more highly of one another than do either the X or Z pupils, which corresponds to the scores actually achieved. The teachers feel much the same way about it, but, as in the case of the deportment marks, the per cent of pupils receiving scores showing a definite reputation for helpfulness is about the same in Y as in Z. When teachers' and pupils' opinions are combined, as in our total reputation score, the differences between the schools in the matter of reputation become negligible — that is, in their own eyes, although wide

TABLE XXX

SUMMARY OF POPULATION DIFFERENCES *

	AVERAGE			VARIABILITY		
	X	Y	Z	X	Y	Z
Age	2½	2½	1	3	1	2
Intelligence	1	2	3	1	3	2
Emotional stability	1	2	3	3	1½	1½
Physical ability	2	3	1	3	1	2
Physical condition	2	3	1	3	1½	1½
Burdick	1	2	3	3	1½	1½
Sims	1	2	–	2	1	—
Occupation	1	2	3	–	—	—
Teachers' marks	1	2	3	2	3	1
Deportment	1	2½	2½	3	1	2
Local reputation	2	2	2	3	1	2

* 1 means highest average, or greatest standard deviation.

differences among individuals occur, each school is as coöperative as the others.

Summary of Population Differences. Table XXX gives the relative position of each population in relation to the other two on each of the facts now being discussed, in the matter of both average and variability. The digits refer to position, 1 being the top group.

In age, physical ability, and physical condition, Z stands highest. In intelligence, emotional condition, Burdick, Sims, occupation, and scholastic average, it stands lowest. Y is usually in the middle and is usually the most variable. On the whole, then, X is decidedly the superior group and the most homogeneous externally. Y is in the middle and has a wider spread. Z is distinctly inferior.

POPULATION DIFFERENCES IN SERVICE

We are now ready to study the relation between the aspects of the general environment which we have just been discussing and the coöperative behavior of the children living under these circumstances. Do children who are surrounded with every advantage prove more coöperative or less than children from a more ordinary community or from an underprivileged community?

The Total Service Score. Figure 36 gives the percentage distribution of service scores on a scale which makes easy a comparison with Figures 33, 34, and 35. The shifts in position stand out quite clearly when these figures are compared. It can be seen at once that Y (the broken line) is the most helpful group, with X and Z coming in this order. The difference between Y and X is, in terms of its statistical significance,* almost twice as great as the difference between X and Z. In respect to service tendencies, 71% of Y and 43% of Z reach or surpass X's median score as compared with corresponding overlaps of 10% and 5% in the case of general social and economic level.

As these differences seem rather inexplicable as they stand, it

* Y − X is 4.5 times its SE. X − Z is 2.6 times its SE. See Table XXXIX, Appendix C.

might be well to compare the populations on each of the service tests to see whether any explanation can be found. The details are given, as usual, in Appendix C, Tables XL to XLIV.

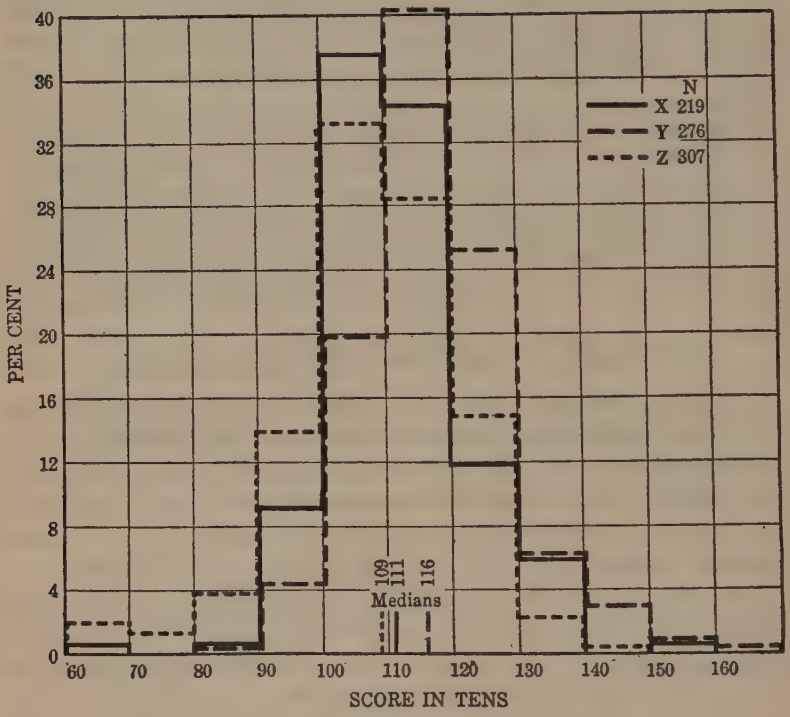


FIGURE 36

PERCENTAGE DISTRIBUTION OF TOTAL SERVICE SCORES BY POPULATION

The facts can best be summarized by emphasizing the contrast between the two tests which set self over against class and the three which introduce the factor of charity.

The Free Choice and Efficiency Coöperation Tests. Figure 37 shows the variation among populations in the median scores on these two "self *versus* class" tests. Evidently, the class as an object of loyal or altruistic behavior appealed less to the X group

than to Y and less even than to Z in the case of the Free Choice test. Z fell down on the Efficiency Coöperation test, which involves continuous effort, tending more to funk the class work and make an effort for self only.

The Money Vote, Kits, and Envelopes. On the Money Vote test, also pictured in Figure 37, population P, a private school, proved most altruistic, and X scored second. On this test, Y, indeed, stands lowest. For the Kits and Envelopes tests, the best basis of comparison is per cent coöperating. The facts appear in

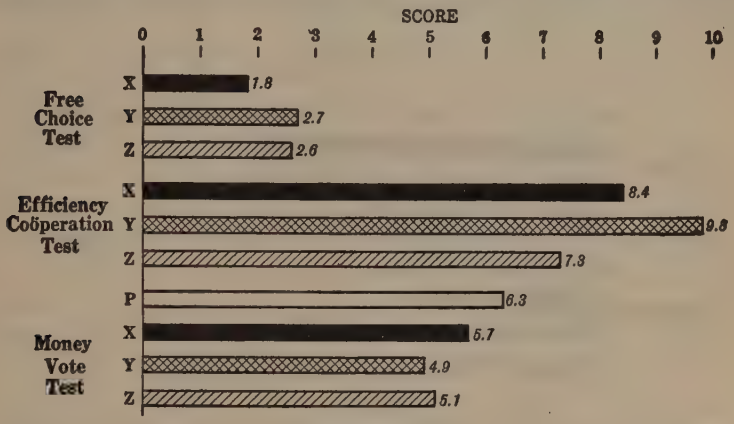


FIGURE 37

POPULATION DIFFERENCES IN MEDIAN SCORES ON THE FREE CHOICE AND EFFICIENCY COÖPERATION TESTS, AND IN MEAN SCORES ON THE MONEY VOTE TEST

Figure 38. The private school is ahead also on the Kits test but falls below Y on the Envelopes. As compared with X and Z, Y maintains its superiority. It is interesting to note, however, that in the *amount* done on the Envelopes test — that is, in the number and quality of jokes, stories, puzzles, and pictures brought in — the children of X who helped at all did better than the children of Y and Z.

Summary of Population Differences in Service Scores. We set out to explain the contrast between population X and Y in the

total service score. X, although significantly higher on a social scale, turns out to be significantly lower on the scale of helpfulness and coöperation. But X is more coöperative than Y on the Money Vote and in the amount of work done on the Envelopes test by those who did anything. The latter might be explained in terms of opportunity — the children of X probably had ready access to more magazines at home than those of Y. The Money Vote test required no *immediate* action but represented the children's ideas of the relative importance of charity, class loyalty, and self-interest. Their votes corresponded to their behavior on the two "class *versus*

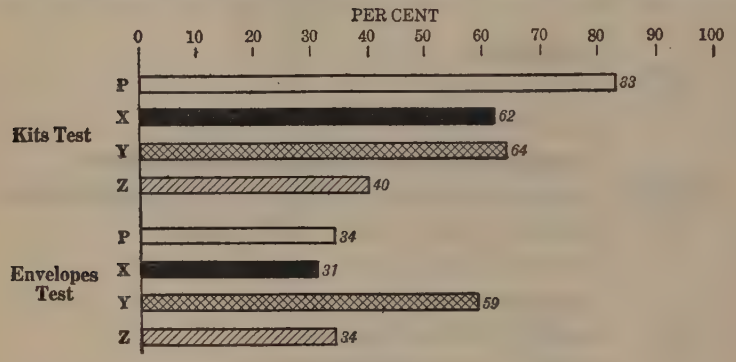


FIGURE 38

Population Differences in Per Cent Coöperating on the Kits and Envelopes Tests — Populations P, X, Y, and Z

self" tests, which showed X to be relatively weak in class loyalty. Although more homogeneous than the other groups in external features of the environment, X is probably more individualistic and less dependent on the classroom as a center of childhood interest. The school does not loom so large in the experience of these more privileged children and does not make so great an appeal to their devotion. Similarly in the case of social service activities, those fostered by the school do not occupy a large place in their imaginations. On the whole, they are both more generous and more coöperative than the children of Z, but less generous and less coöperative than the children of Y.

OCCUPATIONAL DIFFERENCES

So far we have been dealing with general environmental factors differentiating our three populations and have found that, while the lowest group on the social scale is also usually the least generous and coöperative, it is the middle and not the highest social group which is most generous and coöperative. Our problem now is to break up the general environmental factors into such elements as we have been able to measure or observe separately and to see whether the superiority of the middle groups on these independent scales is maintained. The first of these factors is the occupational level of the parents.

The data on the occupations of the fathers were secured from the children in most cases. In populations X and Y they were taken from the Sims Score Card, which calls not only for the name of the occupation but also for several supplementary facts, such as title of position and amount of responsibility. In Z the children supplied the names of the occupations, but this material had to be completed by consulting the city directory.

The occupations were classified somewhat as in our *Studies in Deceit* except that five instead of four groups were used, following the present plan of the Sims Card. These are referred to as levels I to V, I being the highest level and including only professional men and persons holding positions of large responsibility. V contains the lowest forms of unskilled labor.

Figure 39 shows the median total service score for each of these levels for each population.* Evidently it is the highest, not the middle, groups which are most coöperative in each school. The position of X in relation to Y is of especial interest. Each population shows a clear rise in score as the occupational levels go up, but, level for level, Y is superior to X. Indeed the median score of level I in X is below the median score of level V in Y. Occupational differences in Z are not so apparent.†

* See Table XLV, Appendix C, for details.

† As already noted, the occupational data for Z are less reliable than for X and Y.

Comparison with Figure 35, which shows the occupational differences of the three populations, will make it clear that the difference between X and Y in service scores cannot be accounted for by such social facts as may be closely associated with occupation, significant as these apparently are in themselves.

When the service tests are handled separately, the same general occupational differences appear, but they are more definite in the case of the Free Choice and Envelopes tests than with the three

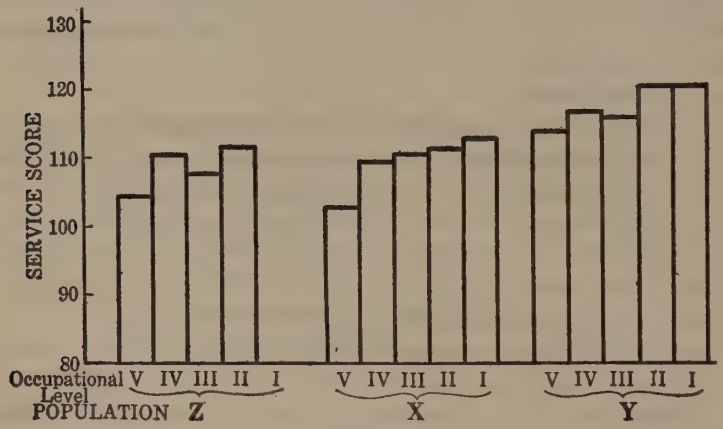


FIGURE 39

Occupational Differences in Median Service Score — Populations X, Y, and Z

others. Tables XLVI and XLVII in Appendix C give the details for all the tests, but Figure 40 makes clear the more conspicuous differences.

In the Z population we ran the medians for the boys and girls separately. The Envelopes test shows no difference between the sexes in the matter of occupational influence. But the others do seem to indicate that whatever background features are associated with the father's occupation affect boys and girls differently. On the Free Choice test fewer and fewer boys give away anything as the occupational levels go down, but the girls maintain their posi-

tion irrespective of level. Similarly, the boys give away less and less as they get lower in the occupational scale, but the girls are not affected. On the Efficiency Coöperation test, requiring endur-

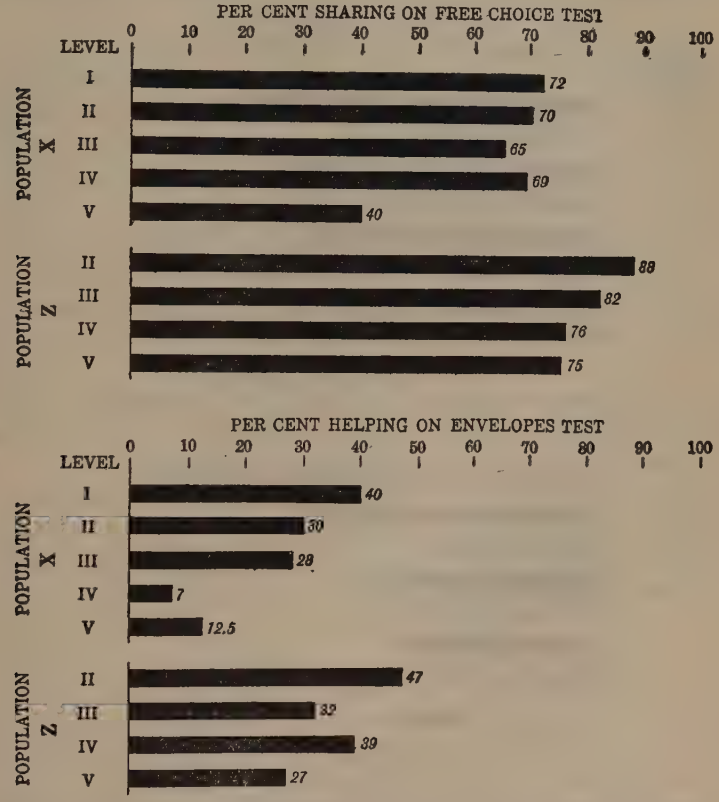


FIGURE 40

Occupational Differences in Two Service Tests — Populations X and Z

ance as well as generosity, the boys hold up and the girls decline as the occupational level lowers, although, level for level, the girls are more coöperative than the boys. On the Kits and Money

Vote, the boys and girls exactly reverse each other. The boys become less and less coöperative and the girls more and more coöperative the lower in the occupational scale they go, with the

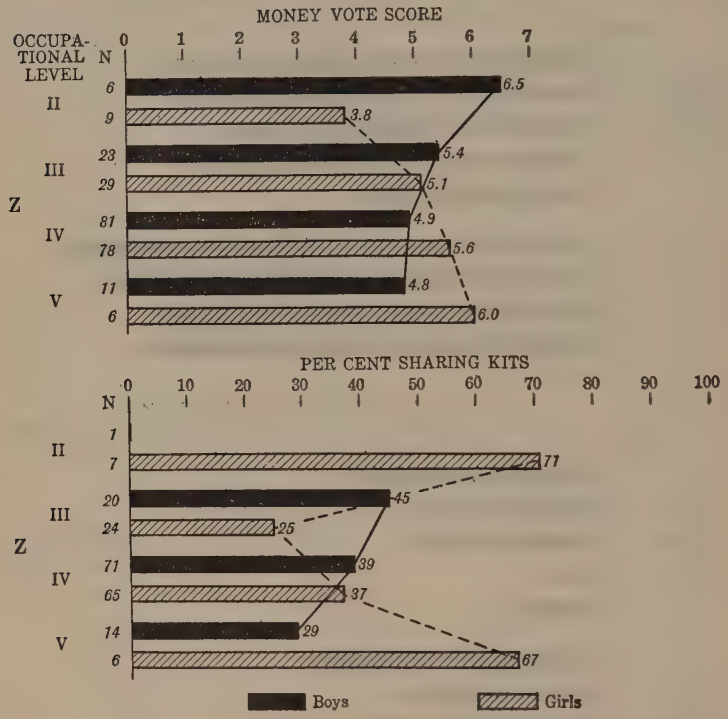


FIGURE 41

CONTRAST BETWEEN BOYS AND GIRLS IN OCCUPATIONAL DIFFERENCES ON THE MONEY VOTE AND KITS TESTS — POPULATION Z

single exception of level II on the Kits test, which has seven girls with a high median (and only one boy). This is such an interesting contrast that we have pictured it in Figure 41. Here it can be seen that the black bars (boys) get shorter and the shaded

bars (girls) longer as the occupational level lowers. Why this should be we do not know.*

ECONOMIC AND CULTURAL DIFFERENCES

The two methods used to get at the general economic and social level of the homes and the cultural background of the children have already been referred to in our discussion of population differences. In this section we shall report the results of these two studies by individuals as well as by populations.

The Burdick Test of Cultural Background. This test, while including items which reveal economic facts, deals for the most part with more subtle aspects of cultural status: range of information about books, magazines, and the theater, conventional etiquette, social practices, attitude of members of the family toward one another, and many other facts of like significance. Obviously, the scores on this test do not depend exclusively on the homes. The correlation between this test and the Sims Score Card, which deals strictly with facts about the homes, is about .50. That is, it would be possible for a child from a poor home to make a relatively high score on the culture test. This is due to the fact that nowadays a child is reached by many cultural influences which are more or less independent of the home, such as the school, the press, the motion picture, children's magazines, churches, and clubs.

Whatever may be the origins of the facts represented in the culture scores, their relation to helpful and coöperative tendencies is of interest, as may be seen from Figure 42, which gives the median service scores for five levels of culture in each population and for the total.† The tendency for service scores to rise as the culture scores rise is clear in all populations, but it is most conspicuous in

* The small numbers at the extremes should warn against an attitude other than that of curiosity.

† The correlations between Burdick culture scores and service scores, corresponding to Figure 42, are as follows: population X, .195; Y, .256; Z, .229; XYZ, .188.

X, which is itself at the top of the Burdick scale. Evidently, cultural differences are of importance at all levels. The fact that the Y population, for reasons seemingly not related to the culture scores, in which it is only a little above Z, is more coöperative than X or Z in some respects, keeps the total picture from showing

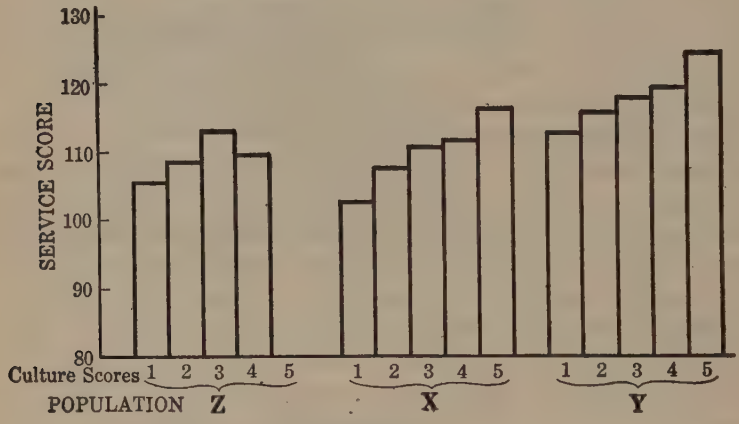


FIGURE 42

CULTURAL DIFFERENCES IN MEDIAN SERVICE SCORES*— POPULATIONS X, Y, AND Z

a clear-cut correlation between service and culture.† But when one population is considered at a time, the results are seen to be quite comparable to those just reported for occupational differences. Such facts as are embodied in the culture scores do not alone account for our population differences, although they are important within each school group. This is more clearly realized if Figure 42 is compared with Figure 33.

---

* The culture scores are grouped here in intervals of 60 as follows: 1 = 36–95, 2 = 96–155, 3 = 156–215, 4 = 216–275, 5 = 276–325. See Table XLVIII, Appendix C, for details.

† These *r*'s are not materially affected when age is partialed out.

We do not have *r*'s between the individual scores on the Burdick and the

The Sims Measure of Socio-economic Level. The relation of the Sims score to service is so slight that it cannot be pictured.* It would not be expected, therefore, that X's superiority in socio-economic level would necessarily give it any advantage in service tendencies.†

Summary of Occupational and Cultural Differences in Service Scores

Cultural differences and the mixture of social facts associated with occupation make more difference in the tendency to be generous and coöperative than do the facts represented in the Sims

separate service tests, but the r's between classroom means in these particulars will show comparable results. These are printed in Table XXXI.

TABLE XXXI

Correlations between 22 Classroom Averages on Burdick and Service Tests

	r
Burdick means and Free Choice medians	.063
Burdick means and Efficiency Coöperation means	.004
Burdick means and Kits means (of givers only)	.192
Burdick medians and Kits per cents who gave anything	.23
Burdick means and Envelopes means (of helpers)	.45
Burdick medians and Envelopes per cents of helpers	.00
Burdick means and Money Vote means	.50

The r's are highest in the case of the Money Vote means, Envelopes means, and Kits. This indicates that charitableness is more closely associated than class loyalty with cultural background.

* The correlations between Sims scores and the total service scores are for X, Y, and XY, respectively, .093, .035, and − .10. The partials, with age constant, are about the same.

† It is an interesting fact that the group means of Sims and service scores correlate .509 in X and .456 in Y. Why should the classrooms differ enough in Sims means to allow so high a correlation? A study of the accompanying scattergram (Figure 43) will reveal a strong relationship between grade and Sims in the case of Y (and for the upper school in X), and we have already

score of socio-economic level.* Within our two populations X and Y, such matters are strongly operative, the higher cultural

pointed out the existence of a correlation between grade and service. The relevant r's are as follows:

CORRELATIONS OF GROUP MEANS IN SERVICE, SIMS, AND GRADE

	POPULATION X	POPULATION Y
Sims and service	.509	.456
Grade and service	− .402	.540
Sims and grade	.000	.691

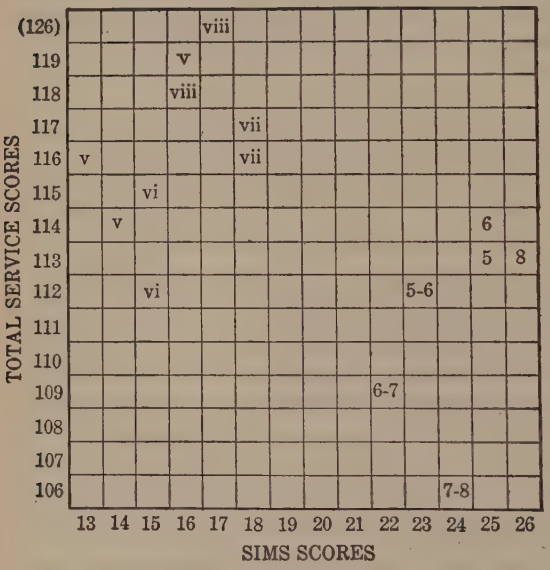


FIGURE 43

SCATTERGRAM OF CLASSROOM MEANS OF SIMS SCORE CARD AND TOTAL SERVICE SCORES FOR POPULATIONS X (ARABIC NUMERALS) AND Y (ROMAN NUMERALS)

When grade is kept constant, the r between Sims and service for Y drops to about .14, which is not far from the individual r of .035 previously noted.

As Sims correlated − .05 with age in Y and − .21 in X, the correlation between grade and Sims needs to be accounted for. It is quite possible that those most likely to drop out of school in the upper grades are not only older but also of lower economic levels, which creates a selection of higher levels. This shows in higher mean scores on the Sims card for the upper grades.

* For the Burdick test the r's reported are: .195, .256, .229, and .188 for X, Y, Z, and XYZ. For Sims the r's are: .093, .035, and − .10 for X, Y, and XY.

levels proving more helpful than the lower. Nevertheless, these cultural differences do not account for the differences found in our populations, for X, although far above Y and Z in social level, is generally less coöperative and helpful than Y and sometimes less than Z.

In addition to the characteristics of the schools themselves, we have left to consider only the national and religious differences between our populations and their possible relation to service. These will be taken up next.

NATIONAL DIFFERENCES

If there are any fundamental biological differences among races, our data are not adapted to their discovery. We do not have distinct racial types in our populations and, even if we did, the social ingredients of behavior are so numerous and effective that we would have to equalize the contrasted races in such matters as culture, ethical standards, economic status, and social experience; and with our few cases this would be quite out of the question. We do not assume that such differences among national groups as we shall report have any biological significance, nor do we deny this possibility. We wish only to remind the reader that the ease with which he is able to conjure up differences in cultural experience among the various national groups should not be mistaken for scientific evidence as to the absence of differences in biological factors.

The details concerning parentage and various service scores are given in Tables XLIX to LI in Appendix C. Let us consider first the total service scores. Figure 44 brings out certain interesting contrasts between the populations. First, it may be seen that the relative position of X, Y, and Z is the same for children of foreign-born and native-born parents and for all combined, namely, Y, X, Z; but is Y, Z, X for those of mixed parentage. Second, the native group is about at the average in each case. In X and Y this might be expected, as the native group is in the great majority; but in Z it constitutes only one-third of the whole. Third, the foreign

group is below the average in all three places, the largest discrepancy occurring in Y. This is shown clearly in the lowest section of the figure, which is for X, Y, and Z combined. Fourth, the admixture of foreign with native born is an advantage in Z, the children of this combination proving decidedly above the average for this population and superior even to those of entirely native

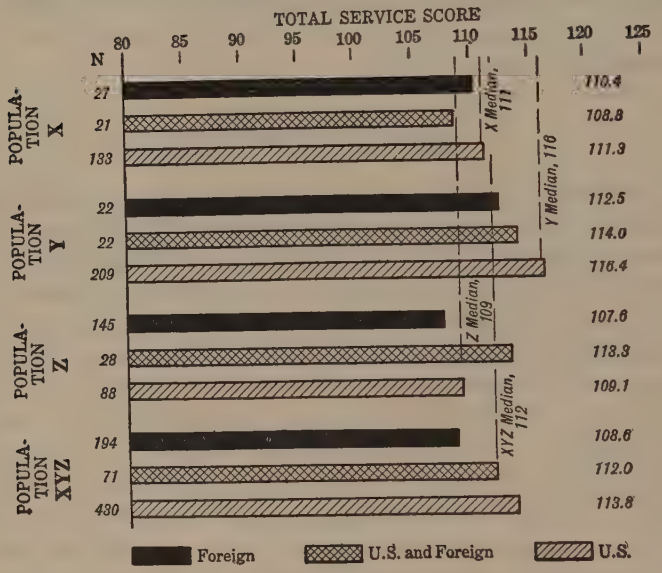


FIGURE 44

COMPARATIVE SERVICE SCORES OF CHILDREN OF NATIVE AND FOREIGN-BORN PARENTS

parentage. In Y also, national intermarriage results in higher service scores, though the result does not equal the average of the population, which is determined by the native born. In X, the mixed group is not only below the average but also below the foreign born; indeed it is 4.5 points below the mixed group of Z. Thus we see that introducing a native element in a predominantly foreign group lowers the level of coöperative and charitable tendencies of the children, but introducing a foreign element into a

predominantly native group raises these tendencies. This is true whether the minority group intermarries or not with the majority group.

Further analysis of the data shows that the least coöperative foreign groups are those whose parents were born in Ireland or

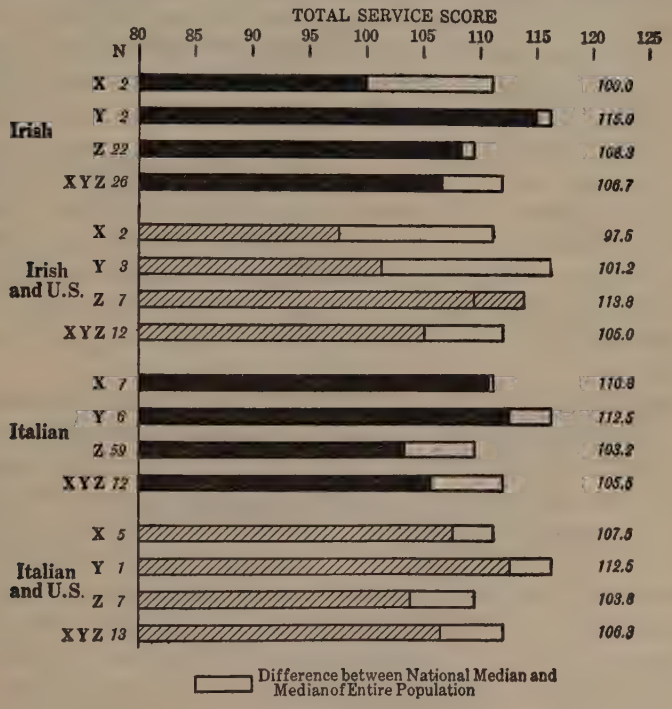


FIGURE 45

MEDIAN SERVICE SCORES OF CHILDREN ONE OR BOTH OF WHOSE PARENTS WERE BORN IN IRELAND OR ITALY

Italy, and the most helpful are those one or both of whose parents were English or Scandinavian. This is true both in the total scores and in the relative position of these groups in *each* of the three populations. This can best be indicated by a chart (Figure 45).

Y retains its superiority over X and Z even in these relatively low groups with the exception of the Irish-Americans, who excel in Z. The extent to which each of these national groups falls below the general average for the population of which it is a part is shown by the unshaded portion of each bar. The Irish-American group in Z is the only one which exceeds the mean of the whole.

Had there been as many of these two national groups in Y as in Z, Y's average would have been lowered; but it should be noted that the Irish, Italian, and Italian-American groups in Y all have higher medians than the average for Z as a whole — higher even than the native group in Z, so that Y's superiority over Z cannot be attributed to the fact that it has a larger proportion of native stock than has Z. The average of Z would of course be raised were the Italian and Irish groups omitted. Of the remainder of the foreign or partly foreign groups, there are 24, or 9%, whose medians are well above the average of the native group and 40 Slavic children, or 15%, whose median is identical with the native group in Z.

National Differences on Separate Tests. Time has not permitted a detailed study of national differences on separate tests, but such facts as we have indicate that the contrast in national groups brought out in the preceding paragraphs is maintained in all but one of the situations. Only 28% of those whose parents were born in Italy, for example, reach the median of those whose parents were born in the United States, in the case of the Efficiency Coöperation test. When all the children of foreign extraction are compared with all of mixed extraction and of native parentage, the story for the separate tests reads just as before except in the case of the Free Choice test, where the foreign children are as coöperative as the native group and those of mixed descent even more than either of the others. The most conspicuous difference in favor of the native children is on the Kits and Envelopes tests combined, where only 26% of those with foreign-born parents reach the median of the native group. One of these tests involves the sharing of objects which might well seem less attractive relatively to the native children than to the foreign, and the other requires

the collection of materials from the homes of the pupils. Consequently, these differences between native and foreign stock may be partly attributable to the relative social status of each. As evidence for this supposition, we cite the facts concerning the

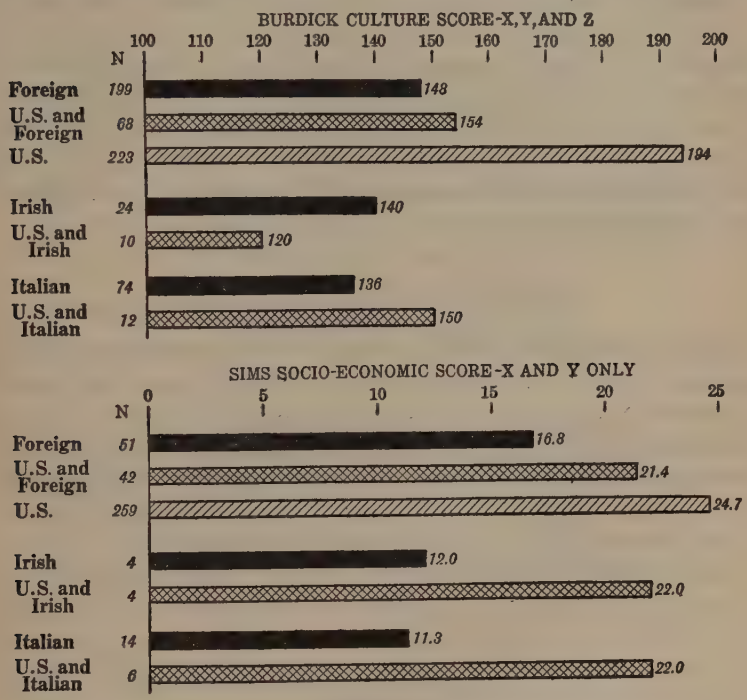


FIGURE 46

National Differences in Median Culture and Socio-economic Scores

proportion of each group falling in each occupational level, as shown in Figure 46. Not only are the culture and economic scores of the foreign group decidedly lower than those of the native group, but the Italian and Irish groups are well below the average of the foreign born; and from Table XLVII in Appendix C it may be seen that on the Kits and Envelopes tests there was a general

tendency for children from the higher social levels to be more coöperative both in numbers and amount done.

As for the Free Choice, it should be borne in mind that this involved the sort of conflict between self and others with respect to which all were on equal terms, so that the failure of the native group to maintain its leadership on this test may properly be attributed either to less interest in the class as an object of appeal or to greater ambition to achieve school success. Their superiority on the Efficiency Choice test, which made the same appeal but involved no self-sacrifice, leads one to believe that the reason for the apparent loss in generosity was rather an excess of ambition than a lack of coöperation.

RELIGIOUS DIFFERENCES

The religious affiliations of parents is a rough index to certain social and cultural heritages, which presumably have their effects on character. Just what these effects are and how they operate has been an age-old problem. Each religious faith claims for itself great moral benefits. It is not our purpose to make general comparisons regarding the moral influences of each religion, but only to present the available data for our three populations.

The number of cases is too small to report each Protestant denomination separately. We shall refer only to the three main groups, Jewish, Catholic, and Protestant. Figure 47, supported by Table LII in Appendix C, gives the important facts for the total service scores.

The relatively low ranking of the Catholic groups is associated with the low socio-economic level of the Catholic children in Z, in which school three-fourths are Catholic and nearly four-fifths are Irish, Italian, and Slavic. Yet in each population the Catholic group runs lowest, as may be seen from Figure 47. About one-fourth of the Catholics reach or exceed the average of the Protestants in X and Y, and about one-third do so in Z. Furthermore, population Y is always superior, and the Protestant group in X only just about equals the Catholic group in Y. Evidently the prepon-

derance of Protestants in Y will not account for its lead in coöperation and charity. The X median would have been a point and a half higher had the Catholics been omitted; but the Y median

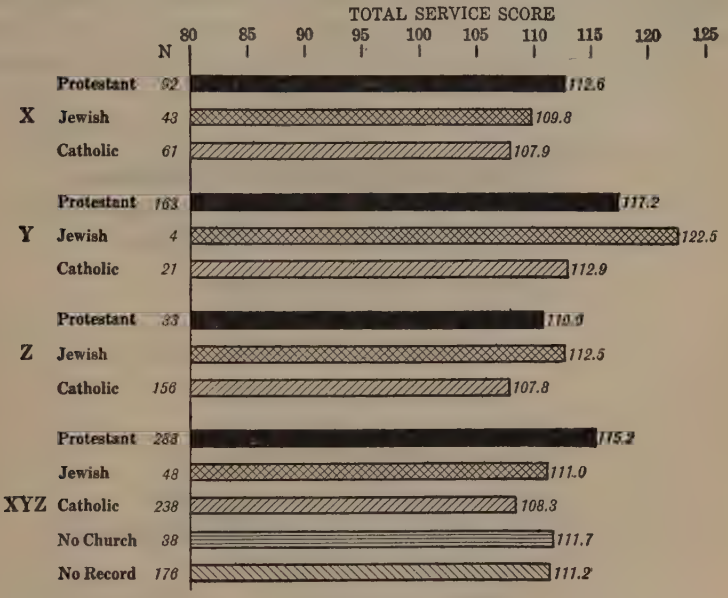


FIGURE 47

COMPARISON OF MEDIAN SERVICE SCORES OF CHILDREN WHOSE PARENTS ARE AFFILIATED WITH THE THREE RELIGIOUS GROUPS — POPULATIONS X, Y, AND Z

would have risen almost as much, maintaining the difference between the two populations.

Expressing these facts in terms of the reliability of these differences, we find that, so far as these particular selections from the respective religious groups are concerned, the differences actually found would occur in nearly the same amount whenever similar situations might be presented. But it should again be recalled that the social selection is rather extreme, as may be seen from Figures 48 and 49, which give the mean culture and Sims scores for

each religious group. The order is generally the same as for the service scores, but the Protestant groups in Y and Z are so much lower on the social scales than in X that, in the combined populations, the socio-cultural order is Jews, Protestants, Catholics, with significant differences between each.* Of course, as has already been noted, the correlation between these social-cultural scales and

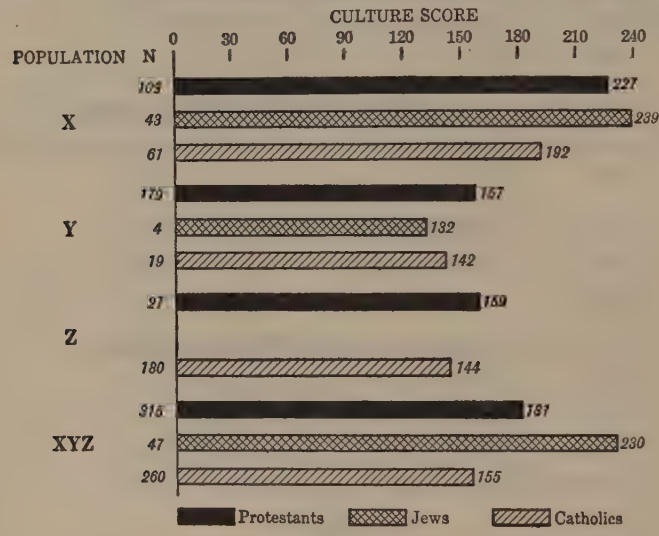


FIGURE 48

Comparison of Mean Culture Scores of Children Whose Parents Are Affiliated with the Three Religious Groups — Populations X, Y, and Z

the service scores is quite low — indeed, in the case of the Sims score, it is zero. These cultural differences, therefore, cannot account for the differences among the religious groups. Neither does intelligence correlate with service to an extent that would make justifiable any allowance for differences in intelligence. In

* The Burdick means differ as follows: Jews minus Protestants, 5.9 σ; Protestants minus Catholics, 5.9 σ. The Sims means differ as follows: Jews minus Protestants, 7.6 σ; Protestants minus Catholics, 2 σ.

the case of the Burdick test we have worked out the service score that might be expected by children whose median Burdick score is the same as that of each of the three religious groups in each population; the results are printed in Table LIV of Appendix C. After giving credit for such differences as might be due to cultural background, the three religious groups still remain apart in service tendencies.

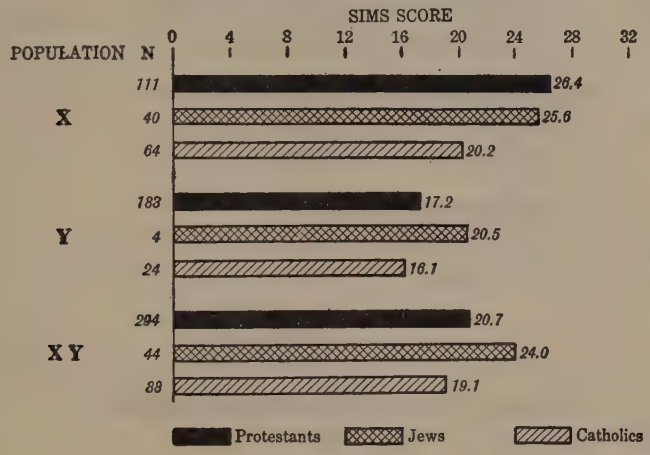


FIGURE 49

Comparison of Mean Sims Scores of Children Whose Parents Are Affiliated with the Three Religious Groups — Populations X and Y

It is of interest to note that, although the culture score of the Catholics rises nearly 50 points between Z and X, the service score remains the same, and although there is practically no rise in their culture score between Z and Y, nevertheless their service score rises 5 points. As in the case of the native and foreign-born groups, the Protestant group, which is superior in high-grade communities in which it constitutes the majority, drops in behavior when it enters a low-grade community as a minority. On the other hand, the Catholic group, which is in the majority in a low-grade community, rises in only one of the high-grade communities which it

enters as a minority — perhaps it is so large a minority in the other that its standards are not so much influenced by the rest of the group as if it were smaller in number. At all events these majority-

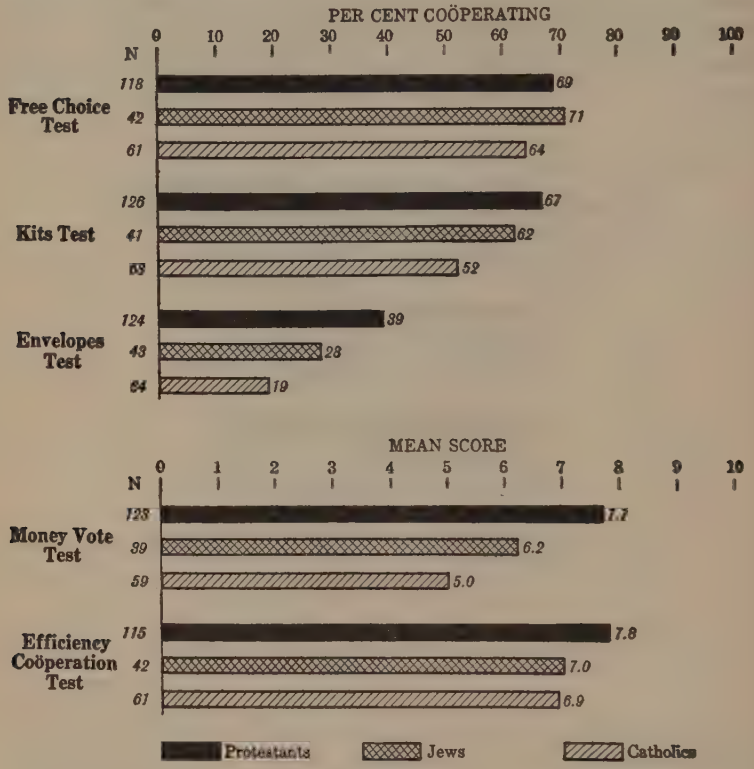


FIGURE 50

Comparison of Scores on Separate Service Tests of Children Whose Parents Are Affiliated with the Three Religious Groups — Population X

minority comparisons correspond to others which we have noted and which will be reported in Chapter XIV.

In order to make a more intensive survey of the three different religious groups and their standing in service, their means on each

of the service tests will be reported for the X school. We selected this school for two reasons: First, each of the three religious groups was well represented here; and second, the Catholic group was composed of a number of national groups rather than of Italians chiefly, as in school Z. The facts are summarized in Figure 50, representing Table LV of Appendix C. It can be seen at once that the relative position of the three groups is maintained throughout, with the exception of the Free Choice test, on which the Jewish group is the most coöperative. The black bar in every other instance is longer than the cross-hatched and the cross-hatched than the shaded.

Summary of National and Religious Differences in Service Scores

Whatever may be involved in the complex of facts called nationality or religion, evidently they are intermingled and influential in behavior. We reported in *Studies in Deceit* that these factors are definitely associated with honesty and dishonesty, although accountable for the most part to the socio-economic and intellectual status. In the tendency to coöperate and be of service we come upon even greater differences, but find less evidence that they are closely tied up with intelligence and the general cultural situation.

CONCLUSIONS

Starting out with large population differences in charitable and coöperative behavior, we have traced back their possible sources in the equally large differences found in occupational distribution, culture scores, socio-economic scores, nationalities, and religion, and have found that, although each of these facts is associated in a positive way with the tendency to be of service, nevertheless, the Y population, which stands far *below* X in most of these environmental factors, is almost invariably *superior* to X in service scores. We shall have to leave open for the time being the question of other possible sources of this superiority, but shall return to it in Chapter XI.

CHAPTER X

THE INFLUENCE OF THE HOME ON THE TENDENCY TO BE OF SERVICE

That character is the product of original nature and environmental influences is a very common, if not indeed trite, observation. The perennial problem before educators is that of separating as far as possible the hereditary factors from the environmental for the purpose of determining the extent to which education is possible. So far this problem has not been solved. It is far too complex for available scientific techniques. Attempts to determine the influence of the home environment as a whole as distinct from school or club environment have met with greater success. When we found, for example, that the tendency to deceive runs in families to about the same degree as intelligence, all we could say was that this resemblance may be accounted for by subtle environmental influences, or by hereditary influences, or by both. About all that we know at present is that intelligence, school ability, and other intellectual attainments, or the lack of them, run in families, and on the moral side that delinquent tendencies also run in families. There is plenty of evidence to show that certain intellectual, emotional, and moral qualities show rather high degrees of family resemblance. But we cannot as yet determine how much of this is biological and how much social.

Even if we cannot distinguish clearly, however, between these two sources of family resemblance, it is still important to determine, if possible, the extent of their combined influence on character. We shall report, therefore, what information we have on how homes influence children in matters of service, although whether this influence is a matter of blood relationship, or precept, or example,

14977

or some unknown admixture of heredity and environment we do not know.

We have three kinds of data that throw light on the way homes may influence the service scores of children: (1) the correlation between siblings (brothers and sisters, treating the family as the unit); (2) a comparison of the service scores of children whose parents coöperated with the Inquiry in securing certain home data with those whose parents did not coöperate; and (3) the correlation between the intelligence of parents and the service scores of the children. Each procedure leads to the same general conclusion: The homes have a definite though small relation to the amount of coöperative and helpful activity shown by the children, and a potent factor in the situation seems to be the example of the parents.

SIBLING RESEMBLANCE IN SERVICE

The conventional method of showing the extent to which any characteristic, such as eye color or intelligence, runs in families is to find the correlation between pairs of siblings. In most families there are at most only two children in grades 5, 6, 7, and 8. When this is the case, the score of one sib is plotted against that of the other, treating the family as a unit having two scores. Both combinations are entered. That is, if one of a pair scores 40 and the other 60, we first put 40 on the X axis and 60 on the Y; then we repeat, putting 60 on the X and 40 on the Y. The resulting scattergram shows the total number of cases rather than of pairs. If there are more than two sibs, all possible combinations are plotted, so that the table shows then the maximum number of combinations.

Now if all children of the same family have exactly the same scores on a test and if the families vary among themselves, the correlation between pairs of siblings will be plus 1.00. If the scores of the siblings are wholly unrelated to each other, the effect would be the same as pairing each sib, not with his brother or sister, but at random, with a child from some other family, and the correlation will be zero. The correlations between siblings on the service tests are presented in Table XXXII. In this table N always means

twice the number of pairs used in the table. It will usually be about the same as the total number of cases. (If there were always only two members in each family, it would be exactly the same as the total number of cases.)

TABLE XXXII

CORRELATION BETWEEN SIBLINGS IN THE SERVICE TESTS

	N	r	r*	r†	$\sqrt{r}$
Efficiency Coöperation	212	.191			
Money Vote	172	.248			
Free Choice + Efficiency	178	.165			.41
Free Choice + Efficiency + Votes	192	.400			.63
Total, omitting Envelopes	122	.367			.61
Free Choice	120		.054(.06)		
Kits	56		.146(− .03)		
Envelopes	64		.396(.35)		
Kits + Envelopes	228			.498	
Total, all tests	202			.420	
Total reputation	206	.294			.54
Intelligence	216	.500			
Burdick	182	.376			

SUPPLEMENTARY r'S

Intelligence and total service (sibs only)	206	.198			
Burdick and total service (sibs only)	199	.100			

* Cases scoring zero were omitted. The biserials between zeros and the rest are given in parentheses. The Envelopes r tends to be spuriously high on account of collusion between brothers and sisters at home.

† These r's are all spuriously large because of the factor of collusion in the Envelopes test. The zeros are included in these plots, but few appear as zeros on account of the low r between tests. When this column and the one preceding are averaged, they fall in line with the r's of the first column.

In the cases of Free Choice, Kits, and Envelopes the correlations include only those whose scores are above zero. In the Envelopes,

for example, the first r is only on those cases who did something and means sibling resemblance in amounts done by those who did anything at all. In the parentheses to the right is another r representing the relation between doing something and doing nothing. Those who have zero (that is, did nothing) are taken as one category and those who did something as the other. The resulting r shows the tendency for a child to do at least a little if his brother or sister does anything at all.

The correlations for the separate tests show that the sibling resemblance in these specific types of service is on the whole very slight. The higher r in the case of the Envelopes test can be readily accounted for on the grounds of collusion. If one member of a family cut pictures and jokes, collected stories, etc., the other member was perhaps stimulated to do likewise. This is supported by the r in parenthesis, which shows that if one did anything there was a tendency for the other to do something also. The amount of collusion, however, was far less than we found to exist in the case of cheating on home work.*

The case of the Kits score is interesting. The r between the amounts given is low, as already noted, but even between giving and not giving there is practically no relation at all.

When combinations of tests are made, the r's between siblings run higher. In fact, they are about as we found for deceit when the factor of intelligence is allowed for. The r's for combined tests seem disproportionately high as compared with those of single tests, but this is partly explained by the low intercorrelations between the separate tests.

Interpretation. What do these correlations mean concerning the influence of the home on the tendency to be of service? Do they represent only environmental influences? It may be argued that, since children of the same family have about the same home environment and since home environments vary from one family to another, all we have here is an indication of this common environment in one family and of variation among families. In *Studies*

* We reported an r of .70 between siblings in deception on the test taken home, in *Studies in Deceit*, Book One, page 232.

in Deceit we tested this hypothesis in various ways. For one thing we selected pairs of siblings from an orphanage where the home (*i.e.*, the institutional) environment had been as constant as one would find for at least half the lives of the children. When the range of environments among homes was thus experimentally restricted, the drop in the correlations between siblings in deceit was by no means what would have been expected mathematically. Unfortunately the present data do not permit of such experimental reduction in range of home difference.

Another test of environmental hypothesis was worked out mathematically by which we showed that, if the home environment is the sole cause of sibling resemblance in deceit, then the correlation between deception and our measures of this environment must be at least the square root of the r between siblings. This reasoning holds good in the present case. In Table XXXII we show the square roots of the sibling r's for the test combinations and the reputation scores. These run from .41 to .63. In Chapter IX we reported correlations between service and our measures of environment which are very much lower than this. Thus environment, as far as we have measured it, will by no means account for all, or anything like all, the resemblance found between siblings in service. Of course, we do not claim to have measured those subtle environmental influences which may be the very factors that are producing the resemblance. Our environmental tests are admittedly crude instruments, and until more precise measures are available the matter cannot be satisfactorily settled.

Another possible explanation for sibling resemblance in character traits is the resemblance found in intelligence. In *Studies in Deceit* a method of determining this was worked out. It involves knowing the relation between intelligence and service, which is .198 (for the sibling cases only), and also the relation between the intelligence of one member of a pair of sibs and the service of the other. This is almost the same. If it were exactly the same, then the same square root relation would hold as in the case of environment. It is so nearly the same that we can say roughly that, in order for intelligence to account for the observed sibling resem-

blance, the correlation between service and intelligence should be at least .60, whereas in fact it is only about .20.

The case stands, then, on about the same ground as it does with deception. The sibling resemblances in service are considerable, though not quite so high as in the case of anatomical traits and intelligence. This resemblance cannot be accounted for on any environmental hypothesis by using the facts available. If it is due to environment, then we must assume the presence of some subtle unmeasured influences. Among such influences is the example of the parents.

PARENTAL COÖPERATION AND THE SERVICE SCORES OF CHILDREN

The second set of data has to do with the service scores of children whose parents coöperated with us in filling out certain information blanks as compared with the scores of children whose parents did not coöperate. The main facts are as follows:

In population P, of the children whose parents coöperated by filling out the home-functioning record 58% voted that they would work for a class prize in a spelling contest and 42% voted that they would work for self. On the other hand, of the children whose parents did not coöperate 51% voted to work for class and 49% voted to work for self. Of all the children who voted to work for class, 65% had parents who coöperated; of those who worked for self, 35% had parents who coöperated. The mean Envelopes score of the children with coöperating parents was 16.8 as against a score of 11.6 for children of noncoöperating parents. The biserial r between the Envelopes test scores of children of coöperating and noncoöperating parents is .43. Of children of coöperating parents, 78% gave something on the Kits, against 72% of the noncoöperating parents. The Kits test and Self-or-Class test were taken in school without the knowledge of the parents.

In populations X and Y 218 children came from homes where parents coöperated in filling out the information blanks. These children have a mean total service score of 113. Two hundred

eighty children came from homes where the parents did not cooperate. These have a mean service score of 111.5. The difference between these means is 1.5. The chances are about ten to one that this difference would always be greater than zero.

Still further evidence that children of coöperating parents are more charitable and helpful than those of noncoöperating parents is secured from the results of a test which the parents took. In addition to asking the parents of X and Y for informational data, we imposed on them a more difficult task — that of actually taking a test. The children of those parents in X and Y who took the test have a mean total service score of 117, against 113.6 for the children of those who did not take the test. Here the difference is considerable.*

These data, although not so convincing as in the case of population P, tend to confirm the hypothesis that coöperating parents (coöperating at least in the tasks we requested of them) send to school children who are more ready to help others than are those of noncoöperating parents. This would seem to point to example as one source of influence in the establishment of habits of service.

PARENTAL INTELLIGENCE AND THE SERVICE SCORES OF CHILDREN

The third set of data to which we have referred consists of measures of the intelligence of parents as compared with the service scores of their children. In connection with our study of social knowledge and opinion we asked the parents to answer certain questions which the children had also in school. Among the forms used was a social-ethical vocabulary test which had been shown to correlate nearly .90 with the CAVI tests of intelligence.† The

* This difference is 2.8 times its SE.

† See our monograph "Testing the Knowledge of Right and Wrong," Monograph No. 1, July, 1927, Religious Education Association, Chicago; also the third volume of this report and the dissertation by Dr. Gladys C. Schwesinger entitled "The Social-Ethical Significance of Vocabulary," *Teachers College Contributions to Education*, No. 211, 1926.

parents' scores on this test correlate with the children's total service scores .10 in the case of both mothers and fathers, showing that there was an extremely slight tendency for the more intelligent parents to have more coöperative children.

CONCLUSION

The strength of the home influence upon making the children helpful may be regarded as only slightly less than the influence which results in resemblance between siblings in intelligence. It is not attributable, however, to such environmental factors as may be measured or indicated by the Burdick score of culture or by the intelligence of the parents. The influence of example is probably of some importance, but the question of the presence of a biological factor among the sources of individual differences in helpfulness and coöperation is still open.

CHAPTER XI

THE SPECIFIC CHARACTER OF SCHOOL GROUPS

SCHOOL MORALE

In *Studies in Deceit* we drew attention to the existence of extreme differences among schools with respect to the tendency to cheat — differences which could not be wholly accounted for by the selection of children but which seemed to be directly related to features of school policy and atmosphere which we called "morale." Our present populations do not permit of a similar analysis in the matter of service tendencies, as we have only one school in each community.*

The differences among the schools which we have reported, however, do not seem to be wholly accounted for by the general background — at least by any factors of background which we have been able to isolate. Population Y remains stubbornly at the top, no matter what environmental facts are considered. It may well be, therefore, that the schools themselves have certain characteristics in which they are distinguishable from one another and which have some definite association with service tendencies. Casual

* The only possible exception to this situation is in population X, where grades 5 and 6 are in one school and 7 and 8 in another. As grade contrasts are involved, a comparison of schools is somewhat unsafe. Reference to Figure 16, however, on page 126, will show that in spite of the tendency in each of the two schools for the upper grade to be higher in service score than the lower, nevertheless, the school containing the two upper grades is lower than the one containing the two lower grades. There is of course the possibility of population selection, as two communities are involved, though they are both high on the social scale.

observation of the classrooms and administrative control of each school would place population X at the top with respect to the factors that make for high morale, with Y second, and Z last. Hence the source of whatever advantage the school in Y has must be found in some less obvious aspect of the situation.

The first of such factors to be noted was the relative length of attendance of the pupils of the three populations. Table XXXIII shows the facts for each age level. It is quite apparent that the

TABLE XXXIII

LENGTH OF ATTENDANCE OF EACH AGE-GROUP OF EACH SCHOOL

AGE	POPULATION					
	X		Y		Z	
	Number	Median	Number	Median	Number	Median
9	1	4.0	2	4.5	1	1.5
10	17	4.1	26	5.4	1	1.5
11	48	5.5	49	6.1	27	5.1
12	83	5.1	53	6.7	53	3.7
13	50	6.4	68	7.4	79	2.8
14	25	4.2	46	8.2	101	2.4
15	9	5.5	26	9.0	36	2.8
16	4	2.5	15	10.4	18	2.8
17	1	10.5	3	8.5		
18			2	11.5		
r	.08		.48		−.06	

children in Y, age for age, have been longer associated with one another, longer established in the community, than in either X or Z. In Z the transiency is very high, as the average length of attendance of pupils is only about three years. The r's at the bottom of each column tell the same story — for X and Z age and length of attendance are uncorrelated, but in Y the correlation is .48.*

* For the normal group, containing no retarded or accelerated pupils, it is .45 in Y.

But why should length of attendance matter? The correlation between length of attendance and the tendency to be of service is zero in all three schools.* Hence it might be inferred that none of the schools is a factor in changing the service tendencies of the pupils. Certainly we must conclude that each is only a small factor, for, as may be seen from Figure 51, the median scores of those just entering the three schools are as far apart as the median scores of those who have been in attendance for several years. Whatever advantage Y has, it has at the very beginning of any child's career in Y.

The first inference one is tempted to draw is that the families which move into Y are the sort that take hold and help when any opportunity comes along, whereas those which move into X are less coöperative. There is, however, another explanation which is more in harmony with other facts to be reported later, and this is that the association of the children with one another tends to create a situation favorable to joint participation in any group activities. The Y children, growing up together, know one another better than those of X or Z. The newcomers are few, and when they do come, they enter this atmosphere of relative intimacy and coöperation and at once come under its influence.

The corporate spirit of Y is further promoted by Y's relative homogeneity in intelligence. It is far less variable than X, though the difference between Y and Z is not so great.† This makes for easier association among the pupils and aids the process of acquaintance and coöperation.

We had occasion in Chapter VI to note that the contrasting association between age or grade and service in the several populations could not be accounted for by reference to the effect of the schools on individual pupils. We have just seen, in confirmation of this assertion, that the correlation between service and length

* The r's between length of attendance in any one school and the total service score are .022, .091, and .011 for X, Y, and Z. The corresponding r with the accelerated and retarded omitted is .102 for Y.

† The respective sigmas are: X, 9.75; Y, 7.92; Z, 8.25. The difference between the SD of Y and the SD of X is 3.3 times the SE of the difference.

of stay in each school is zero. How, then, shall we interpret the increase of service with age and grade in Y and its decrease in X and Z? As for X, we have two schools to reckon with and shall

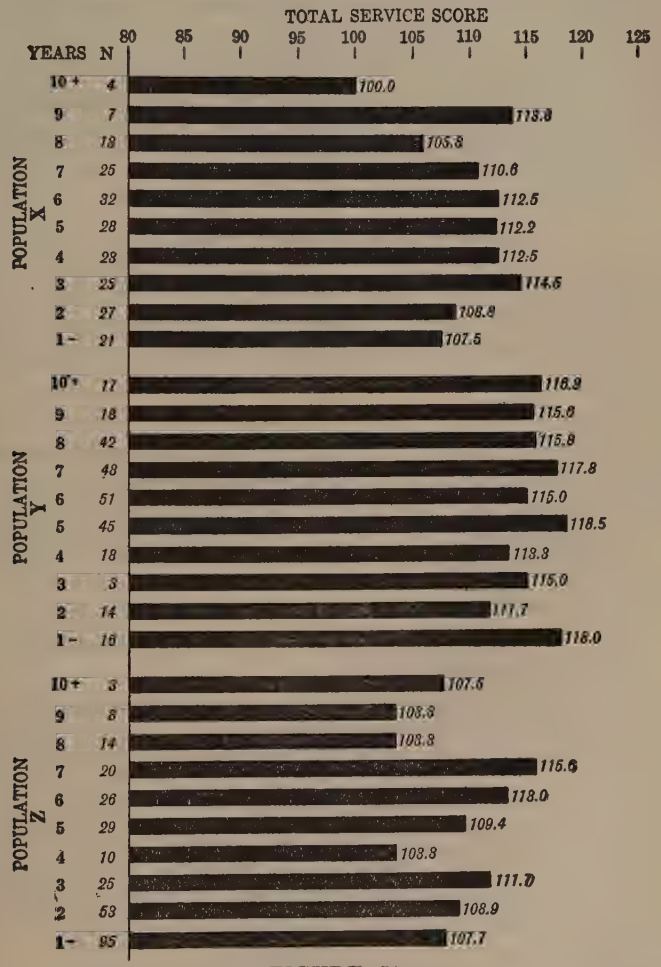


FIGURE 51

MEDIAN SERVICE SCORES ACCORDING TO LENGTH OF ATTENDANCE AT EACH SCHOOL — POPULATIONS X, Y, AND Z

therefore leave it out of consideration for a moment. But in the other populations we find evidence, comparable to that reported in *Studies in Deceit*, for the existence of a *classroom* morale. If this builds up from year to year, as in the case of deceptive tendencies, then this cumulative morale would account for the respective changes with age and grade in Y and Z. We have no evidence regarding such accumulation of morale, as we tested our populations only once, but that genuine differences among classrooms exist is quite apparent. To these we shall now turn.

CLASSROOM MORALE

We have had only one occasion so far to deal with classrooms as units. In reporting the correlation between culture and service we referred to the tendency of classroom culture means to correlate with classroom service means. Such correlation, of course, depended on the existence of classroom differences in the tendency to coöperate on the service tests. If we had thrown all our cases together and then drawn out twenty-two groups at random, each containing 35 children, the average of each group would have varied from the average of the rest only within the limits imposed by the " errors of sampling " — they would have been closely the same. Furthermore, the range of scores found in each small group would have been approximately the same as the range of scores found in the entire population.

Now if instead of drawing twenty-two groups at random we had picked first all retarded children, and then all Catholic children, and then all children born in the United States, the averages of these several groups would have varied far more than we found them to vary by chance, and within each group there would have been greater homogeneity than in the population at large. The reason for these results is the fact that we selected our groups in such a way as to bring together cases which were alike in some factor with which service is closely associated.

As a matter of fact, we have our twenty-two groups already selected for us, and what we wish to know is whether they differ

from one another more than would a series of groups chosen at random. Working backwards the two comparisons which we have just illustrated, we first see whether classrooms are significantly different in service means, and then we compare the range or variability or (conversely) the homogeneity of the classrooms to see whether the pupils resemble one another in each room more than would a group of children selected at random from the entire population.

Comparison of Classroom Means. Inasmuch as grade makes some difference in service scores, we shall have to compare rooms from the same grade which are in all fundamental respects nearly alike. We find available twelve cases in Y and Z.*

In population Y, the rooms are somewhat alike — they resemble one another about as one would expect from a mere chance

*TABLE XXXIV

CERTAIN GROUP DIFFERENCES IN TOTAL SERVICE SCORES

Grade	Room	N	Mean	σ	Difference	$\frac{\text{Dif.}}{\sigma \text{ Dif.}}$
			School Y			
5	N	29	114.40	7.00	O – N	2.0
5	O	33	119.75	13.60		
5	P	34	115.15	8.65	O – P	1.6
6	Q	29	115.25	9.75	Q – R	1.4
6	R	38	111.95	10.05		
8	U	14	118.95	15.55	V – U	1.5
8	V	25	126.10	12.76		
			School Z			
7	C	45	112.05	12.15	C – D	2.8
7	D	46	104.45	14.05		
7	E	42	102.85	12.65	C – E	3.5
8	F	41	110.30	9.35	F – G	4.2
8	G	41	99.55	13.55		

sampling of the population. But in Z there evidently are some factors at work to make one room differ from another of the same grade more than chance would account for.

Comparison of Homogeneity of Rooms. Our second method is too complex to be reported here, but the facts will be found in Tables LVI and LVII of Appendix C. Briefly, when all classes are considered, we find in each population clear evidence of sources of classroom difference far beyond what might be expected by chance selection.*

Resemblance of Boys and Girls within Classrooms. Another procedure for measuring the homogeneity of classrooms consists of dividing each group into two parts and treating the means of the two subgroups as though they were two scores of one individual. The two sets of means — a pair for each class — are then correlated for each population. If each group consisted of a random sampling of the entire population and if each subgroup were a random sampling of the class, the r thus resulting would be zero, for all the factors involved in each mean are the result of chance. But if the members of a class tended to resemble one another in the matter under consideration more than they resembled the members of other classes, the two scores for each class would be more alike than chance would account for, and the correlation would rise above zero in proportion to the homogeneity of the class.

We tried out this procedure; only instead of dividing each class into two groups by making two random selections of individuals, we took all the boys for one group and all the girls for the other, since we had no reason to suppose that the differences between boys and girls differed from class to class in any way but by chance.†

* A rough indication of divergence of classroom means from those of a chance sampling may be obtained by dividing the SD of the classroom means by the mean of the SD's of the several means. The result is 1.98 for X, 1.98 for Y, and 3.16 for Z. Anything greater than one is an indication of some factor other than chance.

† This statement should be qualified somewhat. There may be class differences in the extent to which boys and girls resemble or fail to resemble one another. For example, as will be pointed out later, the majority sex is the more coöperative on one of our tests.

This gave us our two scores on each test for each class, one for boys and one for girls. We then arranged these in rank order for each test and figured the rank difference correlations. The correlations for each population are given in Table XXXV in the accompanying footnote.*

Examination of the results shows that in population Z there is considerable classroom homogeneity, less in X, and a still smaller amount in Y. What we have, however, bears out our previous findings in pointing to the existence of some sort of classroom atmosphere, code, or morale which is accompanied by characteristic behavior patterns.

Resemblance of Classmates Chosen at Random. A final check on homogeneity within classrooms was made by taking each member of each room and pairing him with another member chosen at random from the same room. The amount of binding influence tending to make members resemble one another more than they resembled members of other classrooms would be indicated by the size of the resulting *r*'s. For X the coefficient is .178; †for Y, .062;

* TABLE XXXV

RANK DIFFERENCE CORRELATIONS OF THE CLASS SERVICE SCORES OF BOYS WITH THOSE OF GIRLS — POPULATIONS X, Y, AND Z

TEST	POPULATION			
	X	Y	Z	XYZ
Free Choice — Medians	.33	.76	.92	.58
Efficiency Coöperation — Medians	− .06	.26	.60	.65
Money Vote — Medians	.41	.38	.51	.45
Kits — Per cent sharing	.76	.57	.89	.76
Envelopes — Per cent helping	.74	.62	.75	.89
Total service score	.53	− .04	.95	.60
Coöperation reputation				.10

Unless very high, these *rho*'s should not be given much weight on account of the small number of rooms involved in each case and the consequent unreliability of the coefficients.

† X, it will be recalled, consists of two different schools, which would tend to increase the variability of the classrooms.

and for Z, .112. Again, the facts harmonize with those already reported. Y is a relatively homogeneous population, with only slight classroom differences. In X and Z the classrooms vary somewhat among themselves.

We have no data that would throw any clear light on the causes of classroom variation among the groups of Z, but the teachers represented a wider variety of spirit and training than was the case in Y or X and showed less enthusiasm for the school — which was only natural in view of its situation. This would tend to make them less alike in the amount and character of their influence than the teachers in the other populations.

This suggestion concerning the importance of the teacher's attitude finds support in the case of one classroom in Y. There happended to be two sixth-grade classes, an A and a B. One of these had disciplinary problems during the year. The teacher was ill and had to leave in the middle of the year. This class scored lower than the other class of the same grade in each of the four classroom tests, but scored higher on the test done at home. Furthermore, it was far less homogeneous than the other, the SD of the Efficiency Coöperation test being almost twice that of its neighbor although its mean was two points lower.* Thus this class, lacking possibly the positive influence present in the other, lost its group characteristics and its behavior pattern and tended to disintegrate.

CONCLUSIONS CONCERNING SCHOOL AND CLASSROOM MORALE

The number of classrooms is not adequate to warrant our drawing any detailed comparisons between the effect of morale on honesty and its effect on the tendency to be of service. As nearly as we can judge, the effect in the case of coöperative and charitable behavior is less than in the case of honest and dishonest behavior;

* The distribution is somewhat bimodal — a possible effect of the failure in discipline — the class tending to divide into two opposing groups, the lawless and the law-abiding.

but it is real and important, particularly in population Z, where we find the children becoming less coöperative as they get older and as they enter higher grades.

As for school morale, we have no direct evidence of its existence and must set over against whatever indirect evidence we have the challenging fact that there is no relation between the length of membership in a school and the tendency to be of service. But certain facts about Y, which are true of both school and community, throw some light on the superiority of Y in charity and coöperation. We find that the children of Y are not only more homogeneous in general intellectual ability than those of X or Z but are also more permanently located in their village, thus tending to become better acquainted and building up a basis for community code or practice on a higher level than in the case of X or Z. Of course, were Y not in general a wholesome community, such intimacy as we have indicated would have worked the other way and made the children less rather than more social.

CHAPTER XII

OUT-OF-SCHOOL EXPERIENCE AS A FACTOR IN SERVICE

That general environmental features are closely associated with standards and modes of conduct is well recognized. In the preceding chapters we have shown not only how great are the differences between communities in coöperative tendencies but also to what extent these differences are due to particular aspects of social experience, such as the school, the home, and the religious and national background out of which the family has come. None of these influences seems to be paramount — indeed, none seems to be very great. Possibly the greatest so far noted is that of the religious and national background, but this is intermingled with cultural factors in such a way as merely to direct our attention to the main fact, already pointed out, that the general environment makes a big difference in the behavior of the children brought up under its influence.

There are certain types of moral influence we have not so far mentioned, which can be treated together as out-of-school experience. We might expect that the most important of these as a possible source of tendencies to be of service would be the Sunday school. In addition there are the numerous clubs for children conducted by religious and nonreligious agencies, which we referred to in *Studies in Deceit* as systems X and Y. Next, there were a few children who had attended summer camps and who might conceivably have been made somewhat different by this experience from the other children of their own communities or who might have been selected for the privilege because of some sort of initial superiority. As a fourth influence, we should note the motion picture. Our facts in

all these particulars are ancillary to the main line of our study and are limited not only by the small number of cases but also by the somewhat rough and ready means we had to employ to get the data.

SUNDAY-SCHOOL ATTENDANCE

Population Y. In population Y we procured from the children the name of their Sunday school and Sunday-school teacher. Armed with this information, we went to the schools concerned and requested the records of attendance. We had previously had good luck in finding such records in population A, but here we were supplied with the facts by only three groups: the Catholic, Dutch Reformed, and Baptist. These included a total of only some 84 children out of the entire population. For the rest we know only the fact but not the amount of Sunday-school attendance.

Length of Attendance. The picture for 66 of the children, for whom we found records extending back some years, is seen in Figure 52. There is seen to be a contrast between the score of those in attendance for four years or less and the score of those in attendance for five years or more. It happens that this particular selection of Sunday-school children (which does not include the Catholic group) is above the average for Y as a whole, but this must not be interpreted to mean that those who go to Sunday school are more coöperative than the average, since all but eleven of the Y children attend Sunday school somewhere* and the average for the population is therefore also the average for the Sunday schools. But it is of some interest that the eleven who go nowhere have a median score a point and a half below the median for the rest.

Regularity of Attendance. The Sunday-school records for all three churches showed regularity of attendance for at least a year. We summarized all the facts available and gave each child a score of 1, 2, 3, or 4, representing the average number of Sundays per month the child had attended. The median service scores of the four

* There are, in addition, four Jewish children, with a score far above the average, who have no religious school to go to in Y.

groups for Y are shown also in Figure 52. Again, regularity of attendance is correlated slightly with the tendency to be of service — the scores rise as the frequency of attendance increases.

Populations X and Z. In populations X and Z the number of Sunday schools represented was too great to justify any attempt to get at the school records; so we depended on the statements of

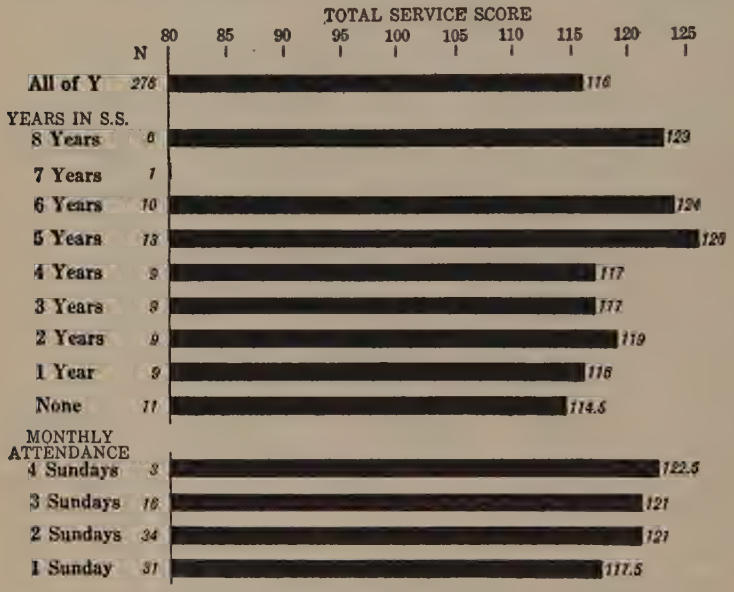


FIGURE 52

SUNDAY-SCHOOL ATTENDANCE IN RELATION TO MEDIAN SERVICE SCORES FOR SELECTED GROUPS IN POPULATION Y

the children. We asked them to answer four questions on a slip of paper passed out at a regular day-school session: first, the name of the Sunday school they attended; second, the name of their Sunday-school teacher; third, the number of Sundays a month they usually went to Sunday school; and fourth, the number of years they had gone to Sunday school. This gave us three sets of data — we shall be sufficiently cautious not to call them " facts " — which,

if they should be facts, might throw some light on Sunday-school influence. We could now separate those who had had Sunday-school training from those who had not had any and could also divide the children into groups with varied amounts of training both in terms of years and in terms of regularity.

Length of Attendance. The contrast between the Sunday-school group as a whole and the children who, according to their own statement, have had no Sunday-school training, is shown in Figure 53 for both X and Z. In X the Sunday-school median is the same as that of the whole population,* and the median for those who have

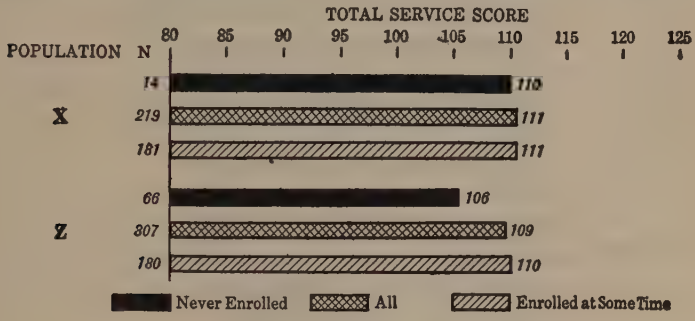


FIGURE 53

SUNDAY-SCHOOL ENROLLMENT AND MEDIAN SERVICE SCORES — POPULATIONS X AND Z

not gone to Sunday school is a point below. In Z the median for those who have had Sunday-school training is a point higher than the median for all pupils, and that of the non-Sunday-school group is three points lower than the median for all.

The case for the non-school group of X is not so good as it first appears, however, for of the 14 children concerned 8 are Jews, who probably go to a religious school on some other day in the week. Their median is 118; and if they are taken out of the non-school group of X, the median of the latter (5 cases only) drops to 103, or 8 points below the median for the school group.

* Including those for whom we had no record of attendance.

When we consider years of attendance, as shown in Figure 54, we find a sharp decline in X on the part of those who have been eight

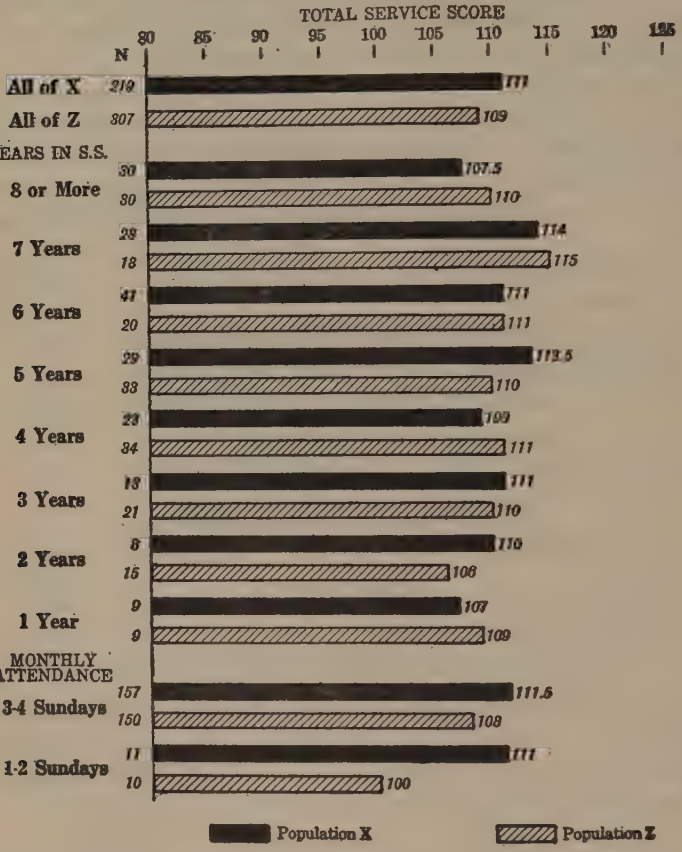


FIGURE 54

SUNDAY-SCHOOL ATTENDANCE IN RELATION TO MEDIAN SERVICE SCORES — POPULATIONS X AND Z

or more years in attendance. Omitting these, the others show a slight tendency for service scores to increase as length of attendance increases. As we saw in our study of day-school attendance, this

relation cannot be attributed to age, as in X the correlation between age and service is negative. Comparison with Figure 51 will also show that the Sunday school, up to seven years of attendance, seems to be of greater positive influence on the X children than the day school, if we consider merely length of association. Unfortunately we cannot compare the day school and Sunday school in the matter of classroom influence.

We cannot account for the falling off in service scores of the 30 children who report a Sunday-school connection lasting for eight or more years. They appear to be in all respects like the rest of the population in age, grade, intelligence, nationality, religion, and social position.

School Z shows a picture similar to that of X except that there is only a slight falling off of the children who have been attending school for eight years or more, and also the correlation with years of attendance for the rest is obviously very small. The 61 children of Z for whom we have no record are a point below the average.

Regularity of Attendance. Turning to the consideration of regularity, we find ourselves at once in difficulty, owing to the nature of the evidence. Nearly three-fourths of both X and Z say that they usually go to Sunday school four times a month! We shall not comment on the veracity of these children, having already discussed in *Studies in Deceit* the relation between honesty and Sunday-school attendance; but if we may assume that there is a real difference in the regularity of those who say they go three or four times a month and those who say they go once or twice a month, then the conclusions pictured in Figure 54 are of some importance. The irregular children of X (attending one or two Sundays a month) are only a half point lower than the regular (attending three or four Sundays a month), but the irregular children of Z are about 8 points lower than the regular, on the service scale. Again, however, we must beware of drawing the conclusion that the Sunday schools are responsible for the situation — some children may be irregular for the same reasons that make them less coöperative. For example, four of the five non-Jewish children of X who report no Sunday-school attendance are Catholic, and the

median of the Catholics of X is 107.9, which, to be sure, is about 5 points higher than that of these four Catholics who do not go to Sunday school. The proportion of Catholics among the non-going children of Z is, however, a little less than for Z as a whole, and the median service score of this non-going group is nearly two points lower than the median of the entire Catholic group.

Conclusions concerning Sunday-School Attendance and Service. Are we justified in drawing any conclusions from these meager data? The fact that the results seem to be constant from population to population, even when based on different types of record, indicates that charitable and coöperative tendencies are probably associated with the amount and regularity of Sunday-school attendance. The association is small, and there is one noteworthy exception in population X, where those in attendance for eight years or more are actually below the median of the entire school. Omitting such peculiar cases, the r between attendance and service would probably not be greater than .10. The biserial r between attendance *versus* non-attendance and service for Z is .242.*

SYSTEMS X AND Y AND SERVICE

The plan of moral education which we referred to as system X in *Studies in Deceit* was not in force in X, Y, or Z. This plan consisted, it will be recalled, in the inculcation of the ideals of chivalry and the effort to get the children to practice good deeds by having them keep a daily record. In Chapter I we reported that in one class of population C, in which we tried out the picture-pasting experiment, a somewhat larger proportion of X's than non-X's came to help, and that the X's did a little more than the non-X's. Also the X's who helped had a slightly higher rank in the organization than those who did nothing.

---

* None of the five children of X who do not go to Sunday school reaches the median of X, which gives a biserial r of + 1.00 between attendance and service. This figure is not reliable, however, because of the small number of cases concerned.

In *Studies in Deceit* we reported the results for three populations for boys who had had the Y type of training for a longer or shorter period. This time we shall report for girls as well. In general, we find, the club members, no matter what the organization, are more coöperative than those not members. This is probably due chiefly to the selected nature of the club membership.*

Let us consider, first, the total service scores. The medians of members and non-members for one boys' organization, two girls' organizations, and one mixed group are shown in Figure 55. By comparison with the scores for all boys and all girls (which, on account of the small numbers of club members, may be regarded as equivalent to the scores of non-members) it may be seen that the boy members are $2\frac{1}{2}$ points higher than the boys who do not belong in both populations.† We may call these the Y_1 group. The Y_2 group, consisting of the girls of a similar organization, has 9 members in X and 29 in Y. In each population it is about 3 points above the median for the girls who do not belong.‡

In the Y_3 group, consisting also of girls, we find 21 members in X and none in Y. The median for these is about the same as for the non-members.

---

* With regard to honesty, the boys in two populations are the same as those not members (in one case slightly more honest and in one case slightly less). In two classes of Z, the low-level school, there were 11 boy members. These were on the average far more honest than their classmates. All membership facts were checked against official lists in X, Y, and Z, as they had also been in A and E. The slight discrepancies found in A and E favored the non-Y's. The only discrepancies found in X, Y, and Z related to honors, which the boys for some reason often did not report identically with the official record.

† The difference in means is .56 σ in favor of the members of Y_1 in population X and 1.2 σ in favor of members of Y_1 in population Y. Although the correlation between the Burdick culture scores and service tendencies is not high, it is of some importance to note that the members of Y_1 are of higher cultural level than non-members in each population. The differences in terms of their SE's are .7 σ for X, 2.9 σ for Y, and 2.5 σ for Z — differences which are relatively greater than those found for service scores.

‡ The difference in means is 1.2 σ in favor of the members of Y_2 in population X and .75 σ in favor of members of Y_2 in population Y.

The mixed group is a young people's society, having members only in population Y. These 12 boys and girls have a median of 123.3, which is $7\frac{1}{2}$ points above the rest.*

It is interesting to compare these differences with those found in the reputation of the children as recorded in the total rating for

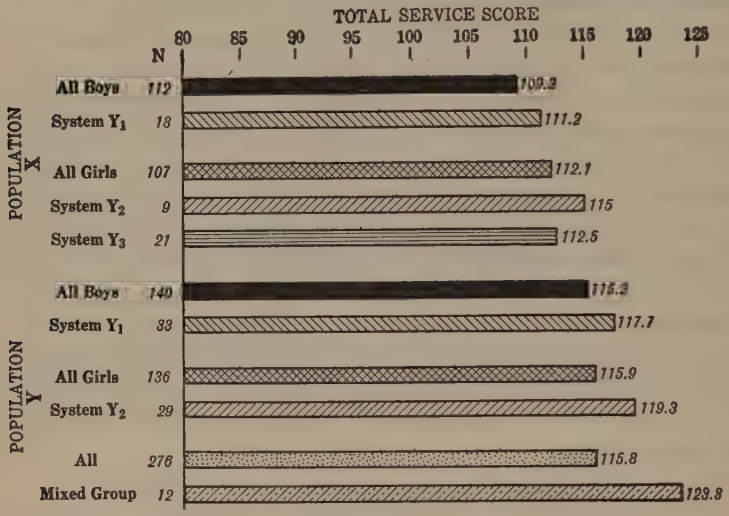


FIGURE 55

MEDIAN SERVICE SCORES OF MEMBERS AND NON-MEMBERS OF ORGANIZATIONS — POPULATIONS X AND Y

service. It is sometimes said that members of these groups are favored by teachers and given credit for what they do not do. In population X, the reverse seems to be the case — the members of Y_1 and Y_2 are both rated as less coöperative and helpful than the respective averages for boys and girls. In Y, however, the boys are rated as average and the girls of Y_2 almost as much above average as the tests actually show them to be. The young people's group stands out as much superior in reputation as in behavior. These facts are shown in Figure 56.

* The difference in means is 1.5 σ in favor of the 12 members of the young people's society.

Ten members of the Y_1 organization in two classes of population Z are as much more helpful than their classmates as are the members of the young people's group in Y.

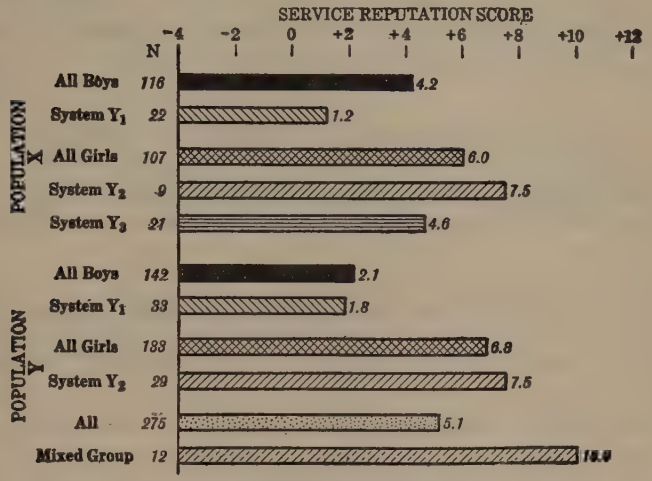


FIGURE 56

COMPARISON OF MEMBERS AND NON-MEMBERS OF ORGANIZATIONS IN REPUTATION FOR SERVICE — POPULATIONS X AND Y

The Y System and the Separate Tests of Service. Of course the advantage of club members is not maintained throughout all the test situations. In population Y the organization we have called Y_1, consisting of boys, is superior on two tests and equal on one and inferior on one. In population X it is superior on two, equal on two, and inferior on one. In the case of the two girls' groups, out of 14 scores (medians or per cents), we find 9 superior and 5 inferior to the general average for girls.

CONCLUSIONS CONCERNING CLUBS AND SERVICE TENDENCIES

The numbers are too small to warrant any comparisons among the members of the several organizations of system Y. But when we compare members and non-members, we find slight differences in favor of the boys' group, somewhat larger differences in favor of

the girls' groups, and a rather extreme difference in favor of the young people's society. The last group contains some very co-operative children, who are either highly selected or well trained. The other groups compare with the total about as do the Sunday-school children, but we have no way of demonstrating any causal relation between membership and service tendencies.

SUMMER CAMPS AND SERVICE SCORES

Some 42 children of population X signified that they went to summer camps. They are just average on all the service tests and on the total.

MOTION PICTURE ATTENDANCE AND SERVICE

The data regarding motion pictures were gathered by means of several questions on the Pupil Data Sheet, which is printed in full in Appendix B. The only point we have followed up is the one about attendance facts — or what we have taken to be facts. The pupils were asked to state how many times a week they attended shows and the answers ranged from none at all to seven or more. We had cases enough of those attending 0, 1, 2, 3, 4, and 5 or more times to make a reasonable comparison of these groups possible.

In the case of population X, as shown in Figure 57, there seems to be a clear break on the total score and on all but one of the several separate tests between those attending up to as much as twice a

* The frequencies for each number of weekly attendances reported are approximately as follows in population X:

Attendance	N
0	45
1	100
2	27
3	14
4	4
5 or more	3

week and those attending more frequently. Indeed, the contrast which is found to be in favor of those attending less frequently

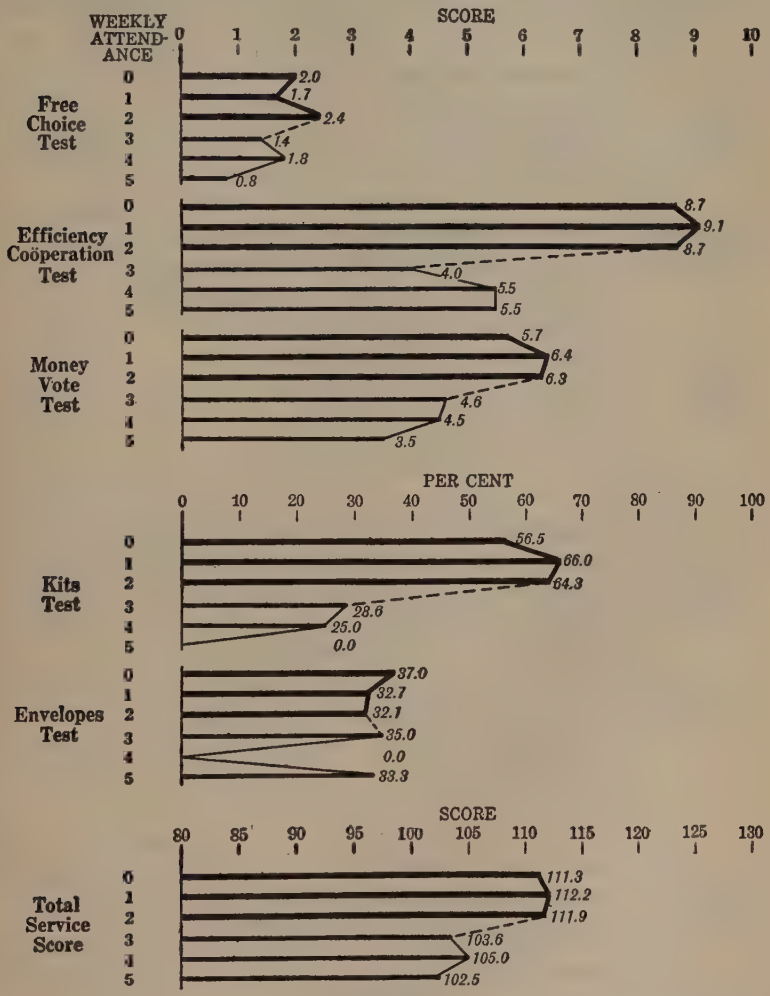


FIGURE 57

MOVIE ATTENDANCE IN RELATION TO SERVICE — POPULATION X *

* See footnote on preceding page.

is somewhat astonishing in almost every case. Population Z (Figure 58), for which we found the facts for the separate tests,

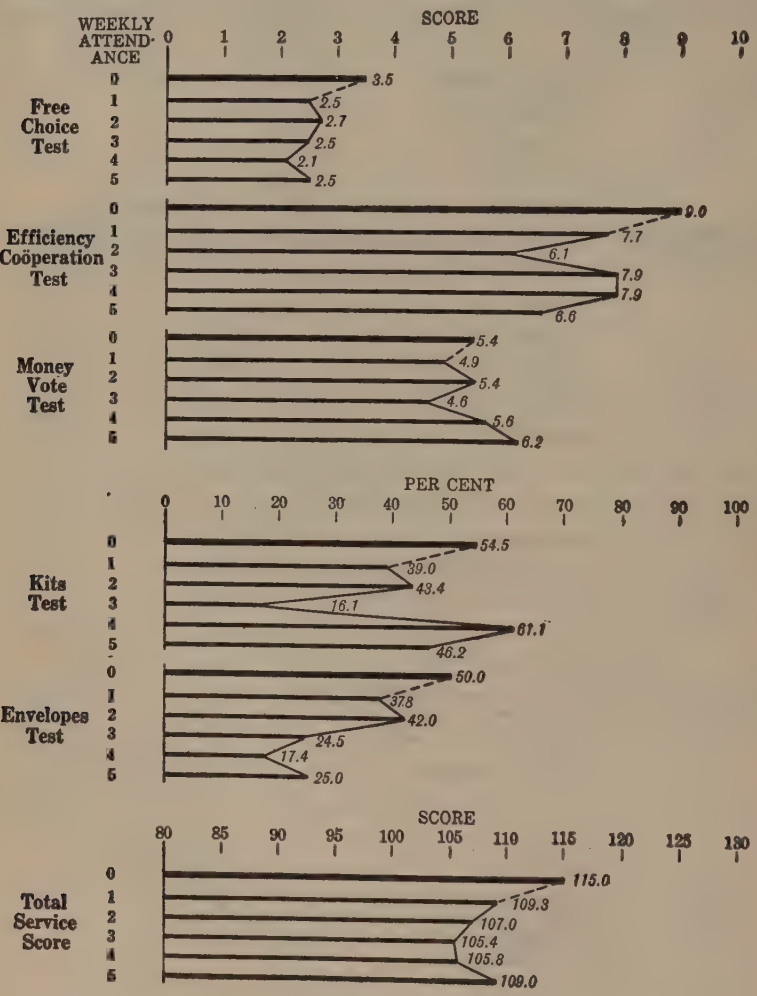


FIGURE 58

MOVIE ATTENDANCE IN RELATION TO SERVICE — POPULATION Z *

* See footnote on next page.

shows a break between those who attend less than once a week or not at all and those who go once a week or more. This break appears on four out of five of the separate tests. We have compared only the total service scores with movie attendance in Y (Figure 59), and here the break occurs between those who attend

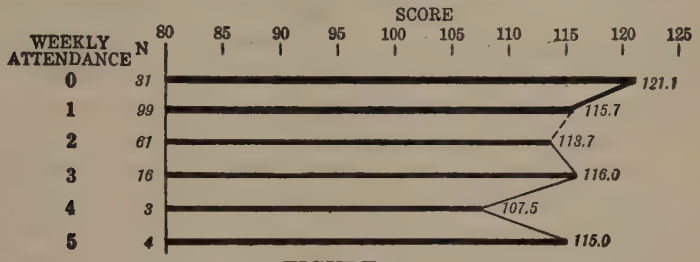


FIGURE 59

MOVIE ATTENDANCE IN RELATION TO TOTAL SERVICE MEDIANS — POPULATION Y

not more than once a week and those who attend more than once a week.

These groupings are of course arbitrary and perhaps to some extent matters of accident. We have, however, figured the statistical significance of the service score differences and find that on the total score they are extreme† and need to be accounted for. A study of the particular cases in X who go more than twice

* The frequencies for each number of weekly attendances reported are approximately as follows in population Z:

ATTENDANCE	N
0	15
1	105
2	70
3	40
4	23
5 or more	15

† In X the difference is 3.8 times its SE. In Y it is 2.4 times its SE. In Z it is 3.0 times its SE.

a week reveals certain facts which in themselves might help to account for the drop in service scores. For example, 55% are Catholic as against 30% for the entire X population. Again, the intelligence median of this group is about 3 points lower than the median for X. Its Woodworth-Mathews neurotic index is a little higher than for all of X, but this might be effect rather than cause. Strangely enough, about half of this group are members of clubs such as system Y, as against about one-third for all of X. No differences in national and economic background are apparent.

One comes often on interesting individual cases in studying over the records. Here, for example, is a seventeen-year-old boy in the eighth grade who reported going to the movies seven times a week. Although getting reasonably good school marks, his reputation is among the lowest of any we have both with teachers and fellow pupils. His deportment record is F. Obviously he is retarded and out of place in his classroom. His parents are Catholic, born in the United States. His score on the service test is rather low, and on some of the tests he did nothing at all.

In the Z population, the exceptionally high scores of the children who do not go to the movies need to be accounted for. We find no economic or intellectual differences to distinguish them, but it is true that there is only one of Italian parentage and one of Irish parentage, whereas 4 are of mixed and 4 of native parentage — the national groups having the better service scores. This phenomenon, therefore, may be a matter of chance and might not be duplicated elsewhere.

Conclusion. It is apparent that other factors than mere attendance at movies need to be taken into consideration, but even after we have noted such concomitants as have been recorded we still find larger differences in service tendencies between the regular and occasional moving picture attendants — between the addicts and the casuals — than it is easy to understand. These differences are about as great, indeed, as among our three distinctly different major populations.

CHAPTER XIII

MISCELLANEOUS FACTORS ASSOCIATED WITH SERVICE TENDENCIES

There remain for consideration these few unclassified facts, suggestibility, cleanliness, sociability, and friendship, concerning which we have reliable data. We have also to report briefly a few studies which afford promising leads for further investigation but which we have been unable to follow up. In some cases further results will be available later. These concern the effect of majorities and teams on service activities, the effect of the repetition of the opportunity to do helpful things, the stability of the coöperation motive, the specific nature of helpfulness and coöperation, and the effect of teaching on coöperation.

SUGGESTIBILITY AND SERVICE

We reported in Chapter I, Table VIII, that those who were better able to resist suggestion of a certain type were also more coöperative. This tentative finding is corroborated for only one population by our more extensive study of this problem, and the association is very slight.

For our test of suggestibility we used portions of the Otis material as published by the C. H. Stoelting Company.* By omitting those items which seemed to us ambiguous, we were able to work in the more useful parts of both forms. Most of the material is of the short-answer type, the subject answering by drawing a line under " yes " or " no " or writing a figure in response to a question

* See also Otis, Margaret, "A Study of Suggestibility in Children," *Archives of Psychology*, No. 70, 1924; also our *Studies in Deceit*, Book One, pp. 277 ff.

which in its mode of presentation suggests a wrong answer. "Do butterflies eat bugs or green leaves?" is a sample of this type. Use was also made of the tendency to keep on giving the same answer in a sequence when the situation called for a change about once in four questions. The suggestive questions were interspersed with irrelevant questions so as not to make the tasks seem too peculiar.

Some of the tests used were based on the suggestive character of large ink blots. We changed the form of these items and introduced them into an art period, having the art teacher administer them as a class exercise. The pupils were provided with mimeographed blanks as follows:

Underline what you think the object looks like, or write in any word which you think the object looks like, thus:

First Sample: dog, boat, engine, cat _____
Second Sample: hat, log, ship, horse ___dog___
1. fish, engine, ship, cat _____
2. boy, hat, boat, dog _____
3. mountain, man, jungle, auto _____
4. cloud, cup, snake, house _____

Holding up a blot that looks like a cat, the examiner said: "What do you think it looks like? I have some others here to show you and I want to know what you think they look like; so I'll pass some papers on which you may write your opinion.

"Now on your papers you see where it says, 'First Sample,' and then the words: dog, boat, engine, cat. The blot I showed you looked like a cat; so the word *cat* is underlined. Now here is another blot. This looks something like a dog, but the word *dog* is not on the list; so it is written on the line at the right. Now when I show you the next four blots, you are to underline what you think the object looks like or write in any other word you think the object looks like. *Look at this,* but don't let on what you think it is. (Picture is exposed for 5 seconds.)

"Did it look like a fish perhaps? Or an engine? Or a ship? Or a cat? Underline what you think it looked like (pause), or write in some other word telling what you think it looked like" (pause); and so on through the set of four pictures.

The pictures in no case looked like any of the objects named by the remotest stretch of imagination, but they could readily seem to be other objects. The score was the number of verbal suggestions taken in preference to making original responses.

The type of resistance to suggestion that is measured by the Otis test is only slightly related to the tendency to be of service, and this relation is doubtless in part due to the common factor of intelligence. The r for Y is .25, but for X and Z it is zero. At the extremes the association is quite visible. The median service score of the 22 most suggestible children of Y is 110, which is 6 points below Y's median, while the median score for the 21 least suggestible is 123.5, which is 7.5 above the median. Every one of the least suggestible is above the median service score of the most suggestible.

CLEANLINESS

In connection with the health examination referred to in Chapter VIII a careful study was made of the cleanliness of the pupils. A

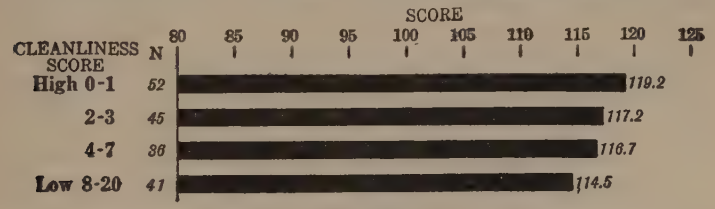


FIGURE 60

MEDIAN SERVICE SCORES FOR DIFFERENT DEGREES OF CLEANLINESS — POPULATION Y

wide range of conditions was found, which seemed to be unrelated to social status. The condition was recorded in the form of a score representing the amount of "old dirt." We grouped the cases in such a way as to compare those whose scores were respectively 0–1, 2–3, 4–7, 8–20, the cleanest having the low scores. The median service scores of each group are shown in Figure 60. Evidently there is an association between keeping oneself clean and being willing to help others, though what the connection is we

are unable to say. Incidentally, those who have 0 scores, that is, who are extremely clean, are three points higher in intelligence rating than the rest,* which may be indicative of the reason for the correspondence.

SOCIABILITY

On our Pupil Data Sheets, following the plan adopted in *Studies in Deceit*, we asked the children to name the playmates or companions whom they liked best. Spaces for four names were allowed. From this information we compiled an objective sociability score consisting of the number of times any child was actually named by others as a playmate or companion.† The range varies from school to school. In X, the boys' scores range from 0 to 11, with the median at 2.7; those of the girls from 0 to 8, with the median at 2.8. In Y, the boys' scores range from 0 to 12, with the median at 2.6, and those of the girls from 0 to 9, with the median at 2.0. In Z, the boys score from 0 to 8, median at 2.6, and the girls, 0 to 13, median at 3.1. Thus Z is the most sociable, with the girls getting more votes than the boys. X comes next, with the boys and girls practically alike. Y comes last, with the boys more sociable than the girls. When boys and girls are combined, the correlations between sociability, as thus measured, and the service scores are in every case zero save in X, where the r for the total score is .10; for the Free Choice test, .12; and for the Kits test (biserial), .23. In each school there is a decided tendency for those receiving the most votes as playmates or companions to fall in the center of the total service scale. It is as though children who stood out in contrast to their fellows were less likely to be chosen as playmates. In population Z only 3, or 16%, of the 19 who scored above 6 in sociability did anything at all on the Envelopes test; that is, 16, or 84%, of them did just what the bulk of the school did.

---

* The difference in means is 3.2 times its SE.

† This, it will be noted, is quite different from asking a child to tell how many friends he has.

FRIENDSHIP

In working out the relation between association of friends and service scores, we used the same data that we have just been discussing but considered only two pupils at a time, the one who filled out the sheet and the one he named first as his friend. Taking these two scores as a pair, corresponding to the paired scores in our study of siblings, we correlated one set with the other. The effect of association between children not in the same class was quite lacking in X and Y but considerable in Z.*

* Table XXXVI gives the *r*'s for both the total service score and for the IER school Xi, or deception score on the three achievement tests. Thus this table may be compared with Table XLIV–A on page 275 of *Studies in Deceit*, Book One.

TABLE XXXVI

RESEMBLANCE OF FRIENDS AND CLASSMATES IN TOTAL SERVICE SCORE AND IER SCHOOL Xi

	SERVICE SCORE						IER SCHOOL Xi						
	X		Y		Z		X		Y		Z		A
	N	*r*	N	*r*	N	*r*	N	*r*	N	*r*	N	*r*	*r*
One friend with one friend in classroom	62	.321	242	.194	340	.347	208	−.023	270	.044	310	.341	.727
One friend with one friend not in classroom . . .	222	.065	206	−.039	136	.351	56	.282	226	−.028	120	.196	.159
One friend with one friend					476	.368					430	.304	.418
One child with a classmate not a friend	204	.178	266	.062	158	.112	186	.053	266	.079	280	.380	.598

The comparable *r*'s from this table are given for population A in the last column. It is apparent that in population A association between classmates meant more than in Z. The influence of friendship apart from the classroom was about the same in the two schools. In X and Y, neither friendship nor classroom association had much influence on either service or deception, which is a surprise.

The facts for mutual friends are given in Table XXXVII, page 238.

Since the classroom association in Z gives rise to a correlation between classmates of only .112 and since friends from different rooms correlate .351, we must attribute the resemblance of .347 between friends who are also classmates in Z primarily to the influence of association outside of classroom. In Y neither friendship nor classroom association seems to influence the children's coöperative and charitable behavior very much. In X, however, in addition to a slight classroom influence, we find that those whose companions are from the same room tend to behave in somewhat the same way on the tests, but that friends from different rooms have no resemblance. In contrast with our findings for deception, then, we may conclude that instead of the classroom being the predominant influence, the association of individual friends is of greater importance in determining service tendencies; but even this is not apparent in Y, and in X it operates only when the friends are drawn from the same room.

Thinking that mutuality in friendship might prove a closer bond than that between two children only one of whom viewed the other as a friend, we ran the correlations for mutual friends in the same classroom, in populations X and Y. The results are not altered for Y but are changed considerably for X. Here, the resemblance between friends from the same classroom is expressed by an r of

TABLE XXXVII

CORRELATION OF MUTUAL FRIENDS OF SAME CLASS IN SERVICE SCORES

TEST	POPULATION X		POPULATION Y	
	N	r	N	r
Total service	102	.567	92	.096
Free Choice	104	− .102	96	.007
Efficiency Coöperation	104	.466	92	.027
Money Vote	62	.200	90	.072
Kits	58	.515	102	.180
Envelopes	104	.277	104	.177
Service reputation	110	.355	104	.133

.567 rather than of .321. The r's for the separate tests correspond. This confirms the impression that association between friends is a real influence in determining their tendency to be of service.

THE EFFECT OF MAJORITIES ON SERVICE

In *Studies in Deceit* * we reported a strong relation between the size of a sex majority and the tendency to deceive in order to win a class prize. Whichever sex had the larger representation tended to cheat more than the other. A similar effect of majorities was evinced in the service tests. We first figured the ratio of the number of boys to the total number of pupils in each class and then the ratio of the median total service score of the boys of each class to the median total service score of the girls. These two sets of ratios were then correlated, yielding for Z an r of .70 but for Y an r of about zero. The Efficiency Coöperation test alone, when similarly treated, showed an r between the two sets of ratios of .76 for population Z. The Free Choice for Z gave an r for similar scores of .22.

Evidently, when the classrooms differ from one another enough to show distinctive characteristics, the fact of being with the majority sex may make a difference in behavior. It need not be supposed, however, that the effect is one of social suggestion, for much of the behavior involved was not obvious at the time. Rather, belonging to the majority sex group in population Z influences the entire situation in such a way as to make it a greater stimulus to the stronger group than to the weaker. In the case of the dishonesty test, the fact of being in the majority made work for the class more worth while. So with the Efficiency Coöperation test or the Free Choice. Possibly on the others the majority sex could more or less dominate the atmosphere of the class, making the minority sex feel somewhat out of adjustment and so less active in almost any situation.

---

* See footnote on page 397 of Book One.

THE EFFECT OF TEAMS ON COÖPERATIVE ACTIVITY

In the case of an eighth-grade group, the classroom coöperation tests were followed the next day by a similar test which, however, placed work for self in opposition to work for one's team rather than to work for one's class. The class was asked to choose two captains, who in turn organized two teams for a "speed contest." The alternating procedure was used as before except that every other unit was devoted to work for one's own team. The Free Choice followed with the usual seven units to be given to one's team or kept for oneself. The results of this latter test offer unmistakable evidence that in this class, at any rate, the motive of working for team was stronger than working for class. On the first day the pupils gave on the average 1.68 units for the class contest. On the second day they gave 3.05 units for the team contest.* This difference, great as it is, is accentuated still further by the fact that in another class, when the regular test was repeated, the work for others fell off rather than rose. In this case it rose 1.37 points — they gave on the average 1.37 more units to their team than to their class. Perhaps the contest was more alluring when opposing groups were self-organized and when all the contestants were visibly present.

THE EFFECT OF THE REPETITION OF SERVICE TESTS

With one group we had occasion to repeat the Efficiency Coöperation and Free Choice tests without any change in the procedure. The reliability of .88 previously reported for the latter test shows that the pupils retain their relative position in each group without much shifting up and down the scale. But in this group there was a significant† falling off in the relative amount of work done

* The difference in the means is over three times its SE.

† The difference between the means was 2.9 times its SE in favor of the first test.

for class as compared with the amount done for self and a falling off in the number of units of work given unselfishly to the class. That is, when the situation was repeated, the interest in self became relatively more prominent — or the interest in others less prominent. The weaker motive lost something of its original appeal.

THE INSTABILITY OF THE MOTIVE OF COÖPERATION

In our discussion of the value of the service tests in Chapter V we mentioned our inability to measure the middle 50% of a large population any more accurately than they could be rated and attributed the difficulty to either the instability or the inactivity of these cases. The demonstration of instability depends upon the possession of a reliable measuring instrument, but the production of a reliable measuring instrument depends upon the stability of the thing that is being measured. If an inch is to be regarded as $\frac{1}{36}$ of a yard, the probability that we shall always get the same length when we divide a yard into 36 equal parts depends on the probability that the yard is itself always the same length. So the building of a test which shall always measure the same thing in the same way can be accomplished only if the thing that the test measures remains constant while the test is repeatedly administered. It is therefore difficult in the extreme to procure reliable tests in the field of character and personality since the behaviors concerned are probably changing in response to subtle aspects of the situation which are not controlled by the test. But this fluctuation, as just pointed out, is hard to demonstrate.

It happens that the Efficiency Coöperation test offers an unusual opportunity for the study of instability in motive. This problem, which is being further investigated, can be only briefly reported on here.

Fluctuation of Coöperative Behavior. If the reader will turn again to the description of the Efficiency Coöperation test on page 68, he will be reminded that the pupils work for twenty-four minutes in units of two minutes each. One unit is done for a prize

for oneself, the next for a prize for the class, the next for oneself, etc. Each unit is in two parts of one minute each. The motive is given just before the first minute is started. The pupils work for sixty seconds and then stop. They are then directed to turn their sheets over and begin again, working for another sixty seconds. They then stop, put this sheet away, and start another sheet, writing either their own name or that of their class at the top, as directed by the examiner, to indicate whether this unit is to be for self or for class. The two successive minutes of work then proceed as before. Thus there are twelve half units for self and twelve for class. Average deviations from his own mean speed score were worked out for every child to show the extent to which his behavior varied from minute to minute around his own general average.* While working for himself these measures of the variability of individual behavior were consistently smaller than when working for the class. The class, as we pointed out in our discussion of this technique, is the weaker motive, resulting in lower efficiency for

* The average deviation of each child's work for self from his mean score for self was divided by his mean score, and in the same way his average deviation in work done in the twelve half units devoted to class from the mean of all his class work was divided by his mean work for class. This made the two sets of scores equivalent in terms comparable to each child's coefficient of variability. The results are shown in Table XXXVIII.

TABLE XXXVIII

INDIVIDUAL VARIABILITY OF WORK FOR SELF AND WORK FOR GROUP

	MEAN	σ	V
Average deviations of individual scores for self around their respective means (MdS)	10.3	5.0	49
Average deviations of individual scores for class around their respective means (MdC)	9.5	4.4	46
MdS ÷ Means for self (approximately VS) . . .	21.7	6.6	30
MdC ÷ Means for class (approximately VC) . .	24.5	12.3	50
$r_{(MdS)(MdC)} = .73$ VC − VS = 2.37			

most pupils, and it has been shown by others that factors causing decreased efficiency, as fatigue, result in increased variability of behavior.*

It can be shown also that, as the test proceeds and fatigue sets in, the wobbling of individuals from minute to minute increases both when working for self and when working for class, which bears out these findings.

Persistence in Coöperative Behavior. In Chapter VII, when we were discussing sex differences, we portrayed the effect of fatigue on the act of writing answers to simple problems in addition. We showed that as the Efficiency Coöperation test proceeded the amount of work done for self in each successive two-minute interval tended to remain about the same or to decrease slightly, but that the work for the class tended always to decrease considerably. Figure 61 will supplement the data previously reported. Here the facts are given for each minute rather than for each two minutes, thus distinguishing the amount of work done on the two sides of each sheet.

The top of each column of the upper chart shows the amount done for its corresponding minute, and the minutes devoted to work for self are drawn in solid lines to distinguish them from the minutes devoted to work for class, which are dashes. The tendency to do the front sheet, when the motive is announced, more rapidly than each back sheet, when the motive has become partly exhausted, is clearly seen. Also the jump upward from each second minute to the work of the new sheet, when the motive is again stated, is seen to be uniformly greater when the new work is for self than when it is for class. Toward the end of the twenty-four minutes, the first minute for self rises to the level reached at the beginning of the test, but the first minute for class drops off conspicuously.

These changes are made clearer in the lower chart, which joins the tops of the columns of the upper chart. The heavy solid line

* See Yule, G. U., "The Increase of Variability with Fatigue," *Industrial Fatigue Research Board Report*, 32, Appendix II; also Weinland, J., "Variability of Performance," *Archives of Psychology*, No. 87, 1927.

and the heavy dashes illustrate what has just been said about what happens during the first minute of each two-minute interval. What happens during the second minute, when the motive has somewhat petered out, is shown by the lighter solid line and dashes. The second minutes for self improve steadily till the last and then fall.

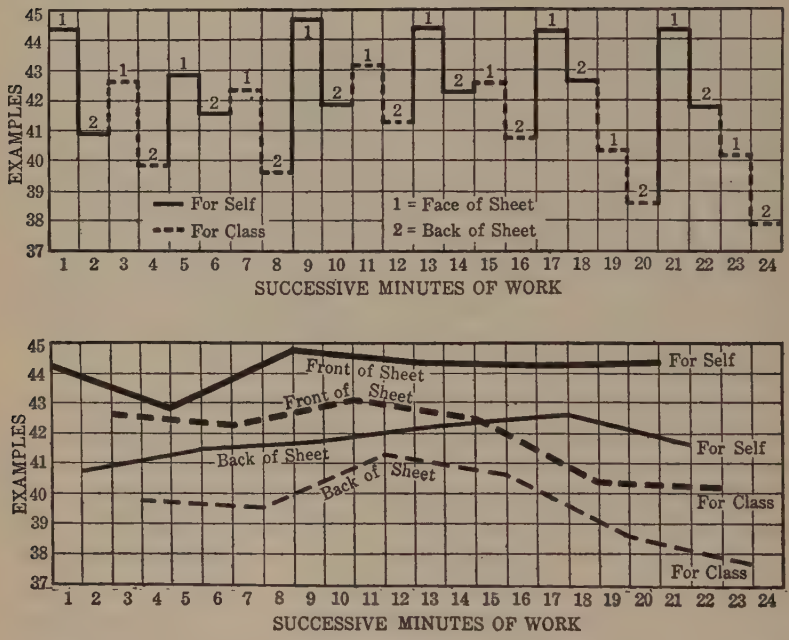


FIGURE 61

Effect of Fatigue on Work Done for Self and for Class — Boys

The second minutes for the class are more fluctuating and begin to fall off after the third time the motive has been given.

The two curves that most nearly agree are those for the second minute for self and the first minute for class — the heavy dashes and the light solid line. At the beginning, the second minute for self is lower than the first minute for class, but it crosses this line

in the fourth work period, as it rises and the other falls. The average amount of work done for self in the second half of all periods devoted to self exactly equals the average amount of work done for class in the first half of all periods devoted to class — that is, the effectivenss of the class motive during its first minutes of operation just equals the effectiveness of the self motive during its second minutes of operation.

All this suggests a possible technique for expressing in exact units the difference between one motive and another.

BIOLOGICAL TYPES AND COÖPERATIVE BEHAVIOR *

Among the measures secured on certain of the Y children by the American Child Health Association were height, weight, girth of hips, and width of shoulders. Two indices were secured from these figures, the weight-height index and the hips-biacromial index. The distributions of these indices were approximately normal. The weight-height index for both boys and girls showed a preponderance of cases toward the tall-thin end of the scale, and the hips-biacromial index showed a preponderance toward the broad shoulders-narrow hips end of the scale — a tendency more conspicuous naturally among the boys, as this body type is characteristically masculine. The presence of preadolescent girls makes the curve for the girls of a more masculine type than would otherwise have been the case.

These measures are of course inadequate for complete study of types. Neither index, however, shows signs of multimodality or grouping of cases around more than one average. If we take the extremes of the distribution, however, as equivalent to contrasting types of body build, we do get significant differences in coöperative behavior as measured by the Efficiency Coöperation test, though not as measured by the Free Choice test. The tall-thin " type " shows less tendency to lie down on the fatiguing task when the self motive is changed to the class motive than does the short-fat

* The study, of which this section is a preliminary report, is being made by Dr. Ruth M. Bakwin.

" type."* There is only one exception, in the case of an extremely fat girl who was also very coöperative.

THE SPECIFICITY OF SERVICE TENDENCIES

Time has not permitted as detailed a study of the specific character of service tendencies as we carried out in our analysis of deception. Such data as we have gathered, however, bear out the conclusions there reached. It will be recalled that the intercorrelations among the service tests † are even lower than those among the deception tests. This may, of course, be interpreted as evidence that they do not measure anything that can be called by one name. Of course, if they all measured exactly the same thing, one test would do for all. So far in the scientific study of mental and social facts no single test has ever been found that would yield the same scores as a large number of tests drawn from the same general field. Almost every intelligence test has a large number of parts dealing with different aspects of what the test maker calls intelligence, and every part has ordinarily several items. The finer the division of parts and items, the lower is their intercorrelation. And quite naturally, for whether one is measuring some factor which is probably a unit, such as " intellect " or " energy," ‡ or is dealing with less clearly defined powers and capacities, all one has available for study is the behavior of the individual in a great variety of situations which themselves are the conditions for the operation of the power that is being measured. All a test does is to arrange a series of such conditions or opportunities and record the resulting behaviors. A total or average behavior is supposed to give a better indication of what the individual can do, by whatever name it may be called, than would any single instance of such

* The difference between the mean coöperation scores of the 25% most tall and thin and the 25% most short and fat is twice its SE for boys and girls combined. Reversing the comparison, the difference in weight-height index between the 25% most coöperative and the 25% least coöperative is 4.2 times its SE.

† See Chapter V.

‡ Spearman, C., *Abilities of Man*, Macmillan, 1927, and Kelley, T., *Cross Roads in the Mind of Man*, Stanford University Press, 1928.

behavior. If such responses were all alike, only one would be needed — their intercorrelation would be + 1.00. If they are not all alike, it shows that one cannot predict from one situation what will be done in another but must observe behavior in both situations. If one wants a general picture, one combines the results. If one wants a more detailed picture, he examines the records for each situation. The problem of what or how many situations to include in a total measure of the individual is a matter of common sense and convenience. One may include situations that closely resemble one another, or situations that vary among themselves even when calling for the same choices between contrasting modes of behavior. The similarity of the situations is determined by observation. The similarity of the responses is determined by their intercorrelation. The lower the *r*'s, the more varied is the response from one situation to another and the less will one situation be able to stand for another.

Prediction of behavior is of two sorts. The first has to do with the probability that when a particular situation is repeated the behavior already noted will also be repeated. This probability is conveniently recorded in the coefficient of reliability. The second sort of prediction has to do with the probability that a type of behavior which has been observed to occur regularly in one type of situation will occur in other situations which are not identical. One may wish to see just to what extent the behavior is affected by a series of controlled changes in the situation. These predictions are possible within the limits indicated by the size of the intercorrelations. Inasmuch as a correlation must be at least .90 for reasonably useful prediction in the case of individual behavior, it can readily be seen that very slight changes in the situation will render prediction uncertain. But if one has a large number of measures of behavior which have a low intercorrelation, if he knows their intercorrelation and the total score of the individual, he can predict the probable total score on a large number of similar situations. The particular situations in which the behavior concerned will appear is, as has just been stated, a function of the reliability of the separate tests, not of their intercorrelation.

We find these propositions borne out by our tests of service just as by our tests of deceit. They have reasonable — though not ideal — reliability, but low intercorrelations. Descriptions of the situations involved show whether they should be called by one name or not, and their intercorrelations show the extent to which behavior on one is related to behavior on another. It is quite clear that the responses are specific, not general, and that we are not able by means of one test or of five tests to divide our population into two groups one of which possesses the trait of helpfulness and the other of which does not.*

THE EFFECT OF TEACHING ON COÖPERATION

The results of the experiment we are about to report must not be regarded as conclusively demonstrating the power of instruction to control behavior. We shall show that in two classrooms in which the teaching was done there was a marked change in behavior in comparison with the change found in two classrooms which were not taught in the interval between the two testings. How soon this difference would wear off is of course not known.

The instruction was given by Mr. Kerr in two schools not used in our main test program. In one, a sixth-grade class was used, and in the other a seventh-grade class. The procedure for the seventh grade on three successive days was as follows:

First Day: Three stories were read to the children. The first was "The Quails," a legend of the Jataka, from *Ethics for Children*, by Ella Lyman Cabot. The second was a "storyized" version of a playlet entitled "Who Is Stronger?" in which the characters are an elephant, a monkey, and an owl, the latter being the "judge." The third was "Hands and Feet," being the story of five sculptors engaged in modeling a clay statue of their king. The latter two stories are found in the Boston School Committee's pamphlet entitled *Citizenship through Character Development* for September, 1927.

* As a check on individual inter-*r*'s we ran the rank difference *r*'s for the classroom means of the separate tests in Y. The average inter-*r* is .16, indicating specificity for groups as well as for individuals.

The first story carries the idea of all persons sharing equally and mutually benefiting thereby, and it also suggests what happens to a group when self-interest breaks up the teamwork. The second story indicates that, while different persons may have different abilities, yet they may be helpful to each other by complementing work. The third story suggests that excusing oneself for not doing his full share is contemptible.

Second Day: The instructor took from his pocket a ball of ordinary red string and said, "I should like to borrow a boy for a few moments." Selecting one from among the many volunteers, he handed him the string and requested him to break it, which the boy easily did. Then the instructor asked for another boy. The second boy was asked to stand about two feet distant from the first. Each extended an index finger, and back and forth, around the extended fingers of the two boys, the instructor stretched strand after strand of the string.

After several strands had been secured, he removed the string, twisted it several times, and handed it to the first boy, saying, "Now, break this."

Then, when the first boy was unable to break the strand, both boys were asked to pull, to see if they could break it. Of course, it was impossible.

Taking the twisted strand away from the boys, the instructor next took a single strand and asked the first boy to break it, which he easily did. The boys then took their seats.

The instructor next requested the classroom teacher to write on the blackboard the sentence, "A chain is no stronger than its weakest link." Children were requested to write in their notebooks these sentences beginning, *viz.*, "A team is no stronger than ——," and "A class is no stronger than ——," with the request that they study the matter and complete the statements in such a way as to convey a meaning similar to that found in the statement about the chain.

Third Day: The instructor first asked for completed statements about the team. The following are samples, as submitted by the children:

"A team is no stronger than its poorest player."

"A team is no stronger than the weaker horse." (!)

"A team is no stronger than its poorest part."

"A team is no stronger than its teamwork."

Discussion revealed that by "poorest player" the children thought of the child who was physically weak and unable therefore to do a full share. They suggested also that a team player might be weak in certain abilities, *e.g.*, passing, basket shooting, foul-tossing, etc., in basket ball. At this point the instructor told of a boy whom he had known, who had been

elected to leadership of a boys' club, largely on account of his ability as a basket-ball player. Within a short time the club began to fall apart, and it shortly became apparent that the elected leader was the cause, primarily because he was a "grand-stand player." "Now," the children were asked, "was he a strong or a weak player?"

At this point someone recalled the story of the quails and applied it to the player being discussed.

Turning to the class idea, completed statements submitted by the children included:

"A class is no stronger than its weakest pupil."

"A class is no stronger than its teamwork."

There was vigorous discussion as to what was meant by "weakest pupil." Some contended that he was the "least able pupil," but finally several children recalled the story of the monkey and the elephant and concluded that the weakest member of a class would be that one who did not do the best he could for the benefit of all the others to the limit of his ability.

In the concluding few minutes the instructor placed a simple scale on the blackboard:

Self 50–50 Others

Children were asked where they thought the ideal pupil should be on such a scale. The majority agreed that they should be on the "Others" end of the scale, at about 60–75. Several advanced the idea that anything one did for the benefit of his class or team was after all of real benefit to himself, too.

In the sixth-grade experiment, the discussion followed the general lines indicated for the seventh grade, but was marked by two differences, *viz:* (1) in discussing "grand-stand playing" half a dozen of the children immediately turned around and pointed at one boy, indicating that that was his style; (2) in placing themselves on the scale, they thought the 50–50 place was just exactly right.

The testing was done quite independently by Dr. Maller. The first test, using the Efficiency Coöperation and Free Choice technique, was given to these two groups and to two control groups the day before the teaching experiment began. The final test was given to the same four groups the day after the experiment was ended. Briefly, the seventh-grade group which received the

teaching increased its mean Efficiency Coöperation score from 7.5 to 9.9 and its mean Free Choice score from 2.0 to 4.5. The corresponding seventh-grade control group, which had received no teaching, in the same period lowered its mean Efficiency Coöperation score from 6.9 to 5.7 and its mean Free Choice from 4.0 to 3.4. In the experimental group there was an increase of 26% (from 63% to 89%) in the number giving something to class on the Free Choice test, and they gave away just about 64% of the total of seven units which they had to give, which corresponds with what the majority said constituted the proper position on any scale of teamwork — 60–75 on a scale of 100. In the control group, on the contrary, there was a decrease of 3% from test to test in the proportion giving something to class.

The results for the sixth grade correspond. The means of the experimental group increased on the two tests respectively from 7.1 to 8.8 and from 2.2 to 3.8, while the mean of the control group decreased from 7.6 to 7.0 and from 3.6 to 3.3. In the case of the experimental group, 4% more (an increase from 76% to 80%) gave away something on the Free Choice test after the teaching than before, but there was a decrease of 2% in the number giving away anything in the control group during the same interval.

It is interesting to note that the amount given away in the experimental group in this grade after the teaching was just 54% of the seven opportunities offered in the test — corresponding to what they had concluded was the proper point on a scale of teamwork, *viz.*, 50–50.

The differences in gain between the experimental and control groups are all fairly reliable* and suggest the possibility of securing at least immediate effects on behavior by means of well-adapted discussion even though the discussion does not concern itself with the conduct included in the test. The pupils were unaware of any collusion between the two experiments, and such connections as were made were probably the result of their own application of the principles of the teaching to the test situation.

* For Efficiency Coöperation test, 6E minus 6C = 2.7 σ, 7E minus 7C = 2.9 σ. For Free Choice test, 6E minus 6C = 2.5 σ, 7E minus 7C = 4.3 σ.

CHAPTER XIV

CHARACTERISTICS OF CONTRASTING GROUPS

COMPARATIVE SIGNIFICANCE OF FACTORS ASSOCIATED WITH SERVICE

In *Studies in Deceit* we examined in some detail the personal history and family relations of a selected number of the most honest and dishonest children in one community. It has not been possible to undertake such extensive case studies in connection with our *Studies in Service*, but we thought it worth while to gather in one place as many facts as time has permitted us to assemble concerning the most and the least coöperative members of our three populations.

We have at times drawn attention to the recognized fact that, even when coefficients of correlation are low, one often finds wide contrasts between the extremes of a scattergram. Inasmuch as the social tendencies measured seem to be in part at least a function of the classroom situation, it occurred to us that it might be worth while to follow the plan adopted in our study of retardation and pull out of each classroom all cases that were either distinctly above or distinctly below the central tendency of the group.*

We endeavored to leave about the middle half in each class out of the picture, since we had already found that they were not well measured on the tests. The boys and girls were selected each with reference to the trend of their own sex. This may be illustrated as follows for grade 6 of school Z:

											Total
N—Boys	2	3	4	5	3	1	3	1			22
N—Girls		2	5	2	5	2	3	2	2	1	24
	90	95	100	105	110	115	120	125	130	135	

Service Score

* This procedure got rid of the influence of heterogeneity and gave us results comparable to what we would have secured if we had used sigma scores.

Thus out of the distribution of the service scores of 22 boys, 5 cases were taken from the bottom and 5 from the top, leaving 12, or 55%. From the girls' distribution, 7 were taken from the bottom and 5 from the top, leaving 50% of the girls in the normal group. In the same way, a few were selected from the bottom and top of each classroom distribution of service scores.

The resulting groups, consisting of the upper and lower quartiles, which we shall label for convenience the selfish (S) and unselfish (U), were then treated as two populations; and means, SD's, and percentages of overlap were figured for each of the more important factors we have been dealing with. This gives a picture of the relative importance of such factors as can be handled in this way.

We shall illustrate the facts graphically and then present the details in tables. Figures 62 and 63 will serve to show how the selfish group overlaps on the unselfish. In Figure 62 an overlap

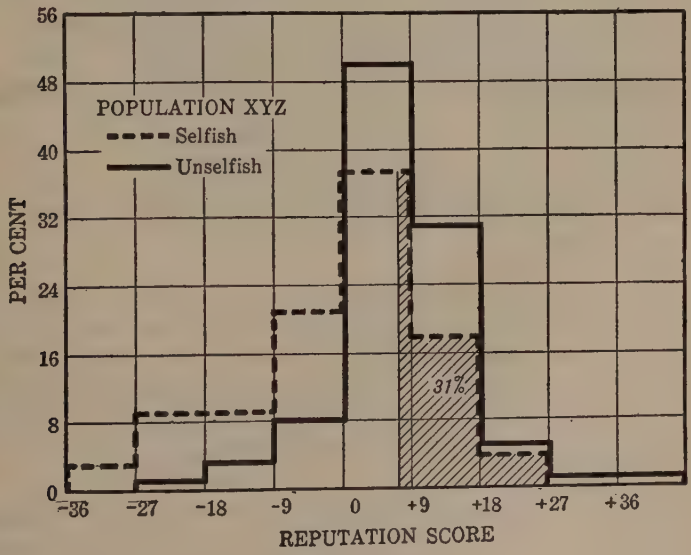


FIGURE 62

PERCENTAGE DISTRIBUTION OF SERVICE REPUTATION SCORES OF SELFISH AND UNSELFISH PUPILS — POPULATION XYZ

of 31% is portrayed; that is, in the matter of reputation for service 31% of the entire selfish group from all three populations reached or exceeded the median reputation score of the unselfish group. In population X the overlap is 36%. In Y it is 30%. In Z it is only 26%.

The overlap for chronological age is shown in Figure 63. Here it is 41%; that is, a little more than half the selfish group fall below the average age of the unselfish group, or, in other words, in age the two groups are somewhat the same.

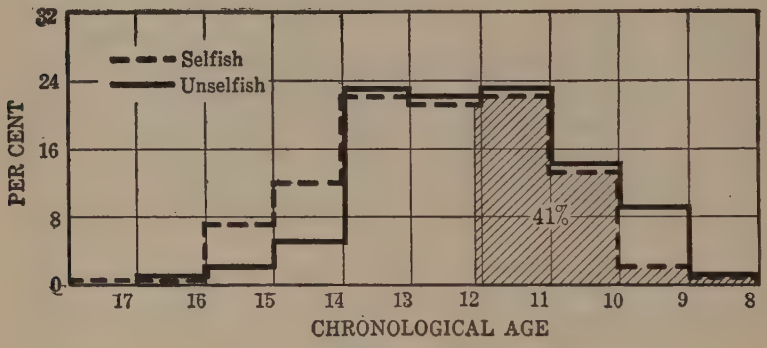


FIGURE 63

PERCENTAGE DISTRIBUTION OF AGES OF SELFISH AND UNSELFISH PUPILS — POPULATION XYZ

Complete distributions for all the factors on which we have compared the two extreme groups are printed in Appendix C. Careful comparison reveals that in amount of overlap, beginning with the least, the order of factors is as follows: (1) reputation, (2) physical condition, (3) cultural background, (4) motion picture attendance, (5) intelligence, (6) emotional stability, (7) age. There is little difference among the last four in amount of overlap.

In addition, we computed the per cent of all children of native-born, foreign-born, and mixed parentage falling in the unselfish and selfish groups and found the significance of the differences between national groups. So also for the three religious groups represented in our populations. From these results, which are

pictured in Figures 64 and 65, it may be seen that, as we noted before, the native group is the most coöperative on the whole, with the mixed group next, and the foreign group last. The pro-

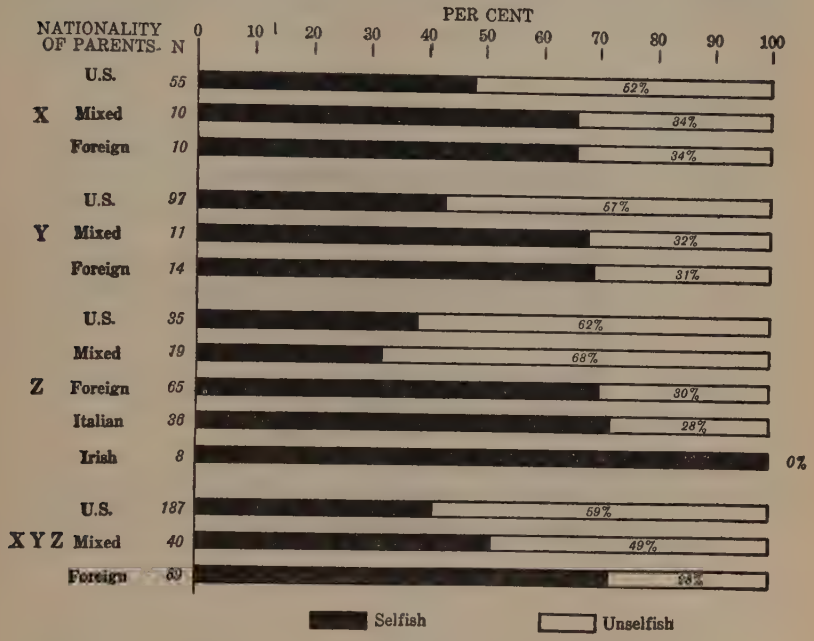


FIGURE 64

PROPORTIONATE REPRESENTATION* OF NATIONAL GROUPS AMONG THE SELFISH AND UNSELFISH PUPILS — POPULATIONS X, Y, AND Z

This figure reads: Of the 55 children of population X in the selfish and unselfish groups whose parents were born in the United States, 52% were among the unselfish and 48% among the selfish; etc.

portion of Italian and Irish groups found in the upper quarter of the service scale is relatively small as compared with the rest. The religious affiliation of parents also shows up clearly as an important factor in distinguishing groups U and S.

* For explanation of the method of figuring the percentages, see footnote to Table XXXIX.

The effect of club membership is treated in the same way. Figure 66 shows the per cent of members and non-members falling in the selfish and unselfish groups in the case of three clubs. These differences are considerable, showing as they do the probability of being found in the U group if one is a member.

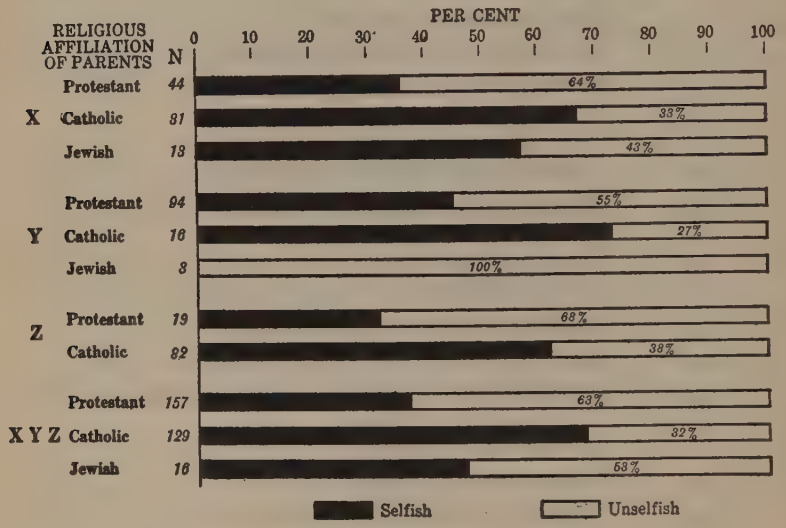


FIGURE 65

PROPORTIONATE REPRESENTATION* OF RELIGIOUS GROUPS AMONG THE SELFISH AND UNSELFISH PUPILS — POPULATIONS X, Y, AND Z

This figure reads: Of the 44 Protestant children of population X in the selfish and unselfish groups, 64% were among the unselfish and 36% among the selfish; etc.

This method of comparison should be checked by the second method used before final conclusions are drawn concerning the relative importance of these several factors in our study of service. This second procedure consists simply in figuring the probability that some difference between the contrasted groups would always occur. This probability is stated in Table XXXIX in terms of

* For explanation of the method of figuring the percentages, see footnote to Table XXXIX.

chances in 10,000. Thus the unselfish group would invariably exceed the selfish group in reputation for unselfishness. They would prove more stable emotionally 9772 times in 10,000 similar experiments, taking all schools together, but only 6915 times in 10,000, which is about 7 times out of 10, or little better than chance, in population Z.

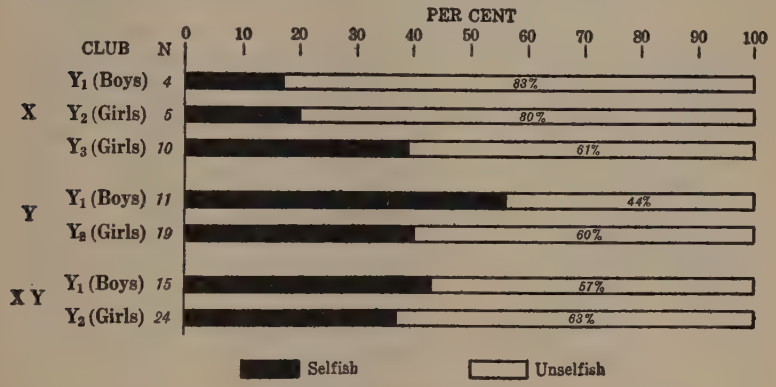


FIGURE 66

PROPORTIONATE REPRESENTATION* OF ORGANIZATIONS (SYSTEM Y) AMONG THE SELFISH AND UNSELFISH PUPILS — POPULATIONS X AND Y

Summary of Comparative Significance of Factors Associated with Service. Doubtless it is rash to attempt any general summary of the amount and significance of the differences between the selfish and unselfish groups with respect to the factors that have just been pictured. It is necessary to keep in mind both the amount and the reliability of each difference shown. But the reliability is chiefly a matter of the number of cases, and for some of our tests and facts there were only a few pupils found at either extreme of the service scale, that is, in the two groups with which we are now dealing. Extreme differences are therefore sometimes associated with great uncertainty as to the probable stability or fun-

* For explanation of the method of figuring the percentages, see footnote to Table XXXIX.

TABLE XXXIX

PROBABILITY OF OCCURRENCE OF DIFFERENCES BETWEEN SELFISH AND UNSELFISH GROUPS

VARIABLE	CHANCES IN 10,000 THAT A DIFFERENCE WILL OCCUR			
	X	Y	Z	XYZ
1. Reputation	9990	9998	10000[1]	10000[3]
2. Cultural background	9990	9452	9996	10000[4]
3. Movie attendance	9821	9965	9554	10000[5]
4. Religion *	9332	8849	9713	9998
5. Age	9192	9032	9974	9986
6. Intelligence	9332	9861	8413	9974
7. Nationality *	6915	9192	10000[2]	9918
8. Emotional stability	9032	7881	6915	9772
9. Physical condition		8849		
10. Clubs *	8849	7257		7881
CLUB, RELIGION, AND NATIONALITY DETAILS *				
Club Y_1 — Boys	9332	6554		
Club Y_2 — Girls	9032	7881		
Club Y_3 — Girls	7580			
Protestants	9641	8413	9554	
Catholics	9641	9713	9772	
Jews	5793			
Native parentage	6179	9192	9332	
Mixed parentage	8413	8849	9554	
Foreign parentage	8413	9332	9981	

[1] 4.5σ [2] 4.3σ [3] 6.7σ [4] 4.2σ [5] 4.0σ

* In the case of these groups the figure given represents the probability value of the significance of the difference between the percentage of the representation of each group found among the selfish and that found among the unselfish. The percentages were weighted so as to allow for the fact that the total number of unselfish did not equal the total number of selfish. Thus there were in population X 41 unselfish, or U's, and 50 selfish, or S's. There were, in all, 44 Protestants found in the U and S groups, of which 26 were U and 18, S. Therefore, to allow for the fact that the proportion of U to S was 41 : 50, we took as our ratio of Protestant U's to Protestant S's $\frac{26}{41} : \frac{18}{50}$. The figures for

damental character of such differences. This is notably true in the case of certain of the clubs.

Making such allowances for these matters as we can, we suggest the following as the order in which the various facts named are associated with service tendencies: reputation, cultural background, religion, movie attendance, intelligence, nationality, clubs, age, emotional and physical condition. In the next chapter we shall attempt a more complete summary, including the factors in which we did not compare the U and S groups.

THE PREVALENCE AND STRENGTH OF UNSELFISHNESS

In two of our tests — Kits and Free Choice — there was a straight conflict between self and others. Furthermore, an immediate decision was demanded. In the first case, the choice was between keeping attractive but not essential objects for oneself and giving them to children in need. In the second case, the choice was between keeping one's work for oneself in order to win a prize and giving it to one's class in order to help the class win a prize. There were a good many children who kept everything for themselves on each test. The percentages in these populations who did not share the contents of the pencil boxes were 38, 36, and 60 in X, Y, and Z, respectively, and 44 for all three populations combined. The percentages who gave nothing to help their class win a prize were 32, 27, and 24 in X, Y, and Z, and 27 for all combined. Evidently, as far as the prevalence of the willingness to share at least a little with others is concerned, we must conclude that from one-half to three-fourths will coöperate. As we saw in the case of the institutional children who were asked to share their ice cream, the proportion varies with the nature of the appeal.

Not only the nature of the stimulus, however, but also the nature

---

the club, religion, and nationality groups in the table represent the standard errors of the differences between such ratios. For Figures 64 to 66 these are translated into percentages.

of the sacrifice must be taken into account. When the appeal is to share attractive objects with children who lack such objects, the proportion of children who decline to coöperate is somewhat larger than when the appeal is to share one's work with one's classmates. The difference is noticeable in X and Y but is extreme in Z, where it might be expected that the contents of the pencil kits, already in hand, would seem far more valuable than a possible, but as yet unwon, prize. Also, the self motive more or less colored the class motive through the identification of self with the group.

In the *amount* shared by the majority who shared anything, we find that the contents of the pencil kits were more attractive than the possible advantage to self to be gained by keeping for self the units of the Free Choice test. There were ten objects in the school kits, and the children could also give away substituted objects. Counting each substitute as equal to $\frac{1}{2}$ an original object, the average per cents of objects given away by those who gave anything in X, Y, and Z were about 20, 25, and 16, and about 20 for all combined. On the Free Choice test there were seven chances all of equal value, yielding a total score of 7. The per cents of this total number actually given to class by those giving anything for X, Y, and Z were 41, 51, and 50, and 47 for all. Again it is clear that, when the children actually shared what they had to give, they were more inclined to keep the contents of the kit box than their units of work. The self motive was the stronger here. In the case of the Free Choice, they divided almost equally between self and class.

A final indication of the relative strength of the self or group motive is found in the Efficiency Coöperation test, where no conflict between self and others was involved. As we reported in connection with our studies of this test, children tend on the whole to work harder and more persistently for themselves than for their class. The per cent of children who did more for their class than for themselves was 28 for X, 47 for Y, 30 for Z, and 35 for all combined. Seven per cent more of the boys than of the girls in X and Y preferred the class to self, and in Z the girls exceeded the boys by the same proportion.

Summary of Discussion of Amount of Unselfishness. In all this we may clearly see that, while the majority of children will share attractive objects with others somewhat meagerly, the majority will share units of work with their classmates about half and half. In working persistently and under fatigue the loyalty to class peters out more rapidly than the interest in self, and it is a definite minority which maintains an average of productivity which is higher for the class than for self. There is no evidence that children are unusually selfish or that they can be divided into selfish and unselfish groups save in the arbitrary fashion we have used in this chapter. Rather is the majority somewhat coöperative, with the rest tapering off gradually up the scale toward extreme unselfishness and gradually down the scale toward extreme selfishness. There is little reason to suppose that, under other conditions discriminating between selfish and unselfish behavior, those at the top in this experiment might not have been found nearer the middle in some other experiment and those at the bottom here been found nearer the middle or the top there.

PART III

Conclusions and Implications

CHAPTER XV

CONCLUSIONS AND IMPLICATIONS OF STUDIES IN SERVICE

The results of our study of charitable and coöperative modes of behavior should be considered in relation to what we set out to do. The general plan which we devised for the guidance of the Inquiry had two immediate goals, as well as several more remote goals. These immediate or primary objects were (1) to determine the practicability of measuring significant conduct trends and (2) to throw as much light as possible on the nature and causes of these trends. The more remote or secondary ends we had in view included the discovery of the way in which these conduct trends were related to one another and to such other aspects of character as social information, opinion, and attitude — in other words, we wished to build, if possible, a test or battery of tests which would measure " character " in the way intelligence tests measure " intelligence " or ability tests measure school progress.

We have published as the first of our sample studies in conduct the volume entitled *Studies in Deceit.* This second volume deals with two more types of behavior, and Book One has been presented as a sample study in the measurement of service tendencies, their nature and causes.

RESULTS OF PRIMARY STUDIES

In the course of our investigation we have tried out a considerable number of techniques and have developed and used five of these in an extensive survey of several hundred children in three communities. These tests may be summarized as follows:

Name	Number of Tests	Number of Responses	Time of Testing in Minutes	Appeal	Desirable Response
Free Choice . .	1	7	10	The class	Sacrificing
Efficiency Coöperation .	1	12	30	The class	Working
Money Vote .	1	5	10	Scaled	Voting money
Kits	1	10	10	Needy children	Sacrificing
Envelopes . .	4	No limit	No limit	Sick children	Working

This battery as a whole has a reliability not less than .80 and probably about .90. Behavior on these tests agrees with the reputation of the subjects about as well as in the case of intelligence tests. The behavior of the middle 50% of the cases tends so to fluctuate that neither the reputations nor the tests can adequately discriminate among them. With respect to that area of behavior which is common to both, however, tests and reputation would agree perfectly, if the measures were free from error.

In order to measure completely the sort of behavior included in these tests, there would be needed thirty-one more like them. Any conclusions, therefore, which have been drawn concerning the characteristics and causes of service tendencies must be regarded as having reference to what children tend to do in the particular types of situation that have been included in the battery. The most serious limitation is in the general background of the tests, four of which involve behavior in school only and even the fifth of which, though done at home, is done in connection with the school. As a sample of the behavior in question, therefore, we have covered less ground than in our *Studies in Deceit*, but have once more demonstrated that a battery which might properly claim to measure any single

type of conduct must include a large number of tests representing a variety of life situations. A beginning in the construction of such a battery has been made in the case of service, and from what has been done one may easily envisage the task as a whole and form some opinion concerning the practicability of the methods which have been employed.

RESULTS OF SECONDARY STUDIES

The second objective concerns the study of the causes and characteristics of service tendencies. In tracing the association of these conduct trends with other facts, we have followed the general practice of research in the mental and social sciences. Selecting contrasting populations, we have administered all our tests to all the children of grades 5 to 8 under standardized conditions. In addition we have gathered all available information about the subjects from records and questionnaires and have, where necessary, administered supplementary tests. This procedure has yielded a vast amount of material, only a portion of which has been statistically treated. The major concomitants of service tendencies have been studied, but there are many details which it would be of interest to investigate with the data now in hand.

Concerning these major concomitants of service, we may conclude as follows:

1. The Relation of Service to Age. In the situations embodied in the tests used for measuring service, no consistent changes were observed. In some environments the children become less coöperative as they get older. In others they become more coöperative as they get older. (For the total service score the r's are $-.203$, $.052$, $-.158$, and $-.18$ in X, Y, Z, and all combined.*)

2. Grade. The factors of age and grade are closely related. Again no consistent differences among the grades in service are found, but such changes as take place seem to be functions of the

* Where r's have been computed for X, Y, and Z combined, they will be presented in this summary as the fourth of the series in each case, the others always appearing in the order X, Y, Z.

total situation. In some situations the children of higher grades are more coöperative. In others they are less coöperative.

3. Grade Retardation and Acceleration. In general the accelerated children are usually more coöperative and the retarded less coöperative than the normally placed. The differences, though of frequent occurrence, are neither consistent nor large, and any observed instances might have been the result of chance. The fact of their repetition, however, suggests that classroom behavior of the sort measured by our tests is partly a function of classroom adjustment.

4. The Relation of Service to Intelligence. As might be inferred from paragraph 3, the brighter children tend slightly to be more coöperative than the normal and dull children. The correlations are low, however (.095, .193, .163, and .161), and by no means show the strength of the association previously reported as found between honesty and intelligence.

5. School Achievement and Deportment. Confirming our suggestion that general classroom adjustment is a factor in the determination of behavior, we found a closer association between school marks and service than between intelligence and service. The relation even here is not strong, however, averaging about .31 for scholastic marks (.266, .239, .346, and .316) and .20 for deportment marks (.154, .209, .223, and .194).

6. Sex Differences. In most tests, and on the total service score, the girls are slightly more coöperative than the boys (1.9 times the SE of the difference) and are regarded as much more cooperative by their teachers and classmates.

7. Physical Condition. No association with service tendencies was found in the case of such aspects of physical condition as were measured by strength and lung capacity (.082, .019, .134, .018). Only a barely perceptible association appeared in the case of such aspects of physical condition as were measured by nutrition and care of teeth (.17), and this might well have been due to such concomitant factors as intelligence and home background.

8. Emotional Condition. On the whole, the relation is about as in the case of general physical condition (.082, .172, .068, and

.152 for the Woodworth-Mathews Sheet). It is noteworthy that the extremely unselfish children only rarely show any evidence of emotional instability.

9. Community Differences. The differences among the three communities studied, though not large on the total service score, are highly reliable and consistent. The mid-population, from the standpoint of intelligence, social and economic status, and almost all measured factors, is the most coöperative; the high group comes second, and the low group last. These differences are difficult to account for, as the next few conclusions will show.

10. Occupational and Cultural Differences. General socio-economic facts make no apparent difference in the amount or prevalence of service tendencies, but such facts as are associated with occupational level and the scores on our tests of culture tend also to be associated with service. As the occupational scale or the Burdick scale ascends, the scores ascend in each population, the *r*'s (.195, .256, .229, and .188) averaging .23 for the three populations. These associations, though small, are consistent.

11. National and Religious Differences. Differences among national and religious groups in such matters as intelligence and social-cultural level do not account for the differences found in tendencies to be of service. In each population the children of English and Scandinavian parentage were more coöperative on the whole than those of native parentage, and the children of native-born parents than those of foreign-born parents. In each population the Italian and Irish groups did less than the rest on the tests. In cases of mixed parentage, the children tended to take on the customs of the majority group.

Of the religious groups the Catholic made the poorest showing and the Protestant the best. The four Jewish children in Y were superior to the children of the two other religious groups, but in X the Jewish children fell a point below the average although their cultural background was much superior.

12. Kinship. The resemblance of siblings in service tendencies is almost as great as in deceit or in intelligence and seems to be partly dependent on biological as well as on social factors.

The *r*'s for separate tests average .207, with about .40 for the total.

13. Example of Parents. Children whose parents take the trouble to fill out blanks at the request of the school are somewhat more coöperative than those whose parents fail to do so.

14. Intelligence of Parents. This is about as closely related to the service tendencies of their children as is the intelligence of the children themselves, *viz.*, an *r* of .10, which is extremely little.

15. School Morale. Schools do not affect the coöperative behavior of children in proportion to the length of their attendance in them. Whatever effect there is operates either at once or by means of a cumulative classroom morale. The effect seems to be in proportion to the intellectual homogeneity of the school and to the *mean* length of attendance of the pupils. These factors would make possible the mutual acquaintanceship which may be a necessary condition for the active coöperation of classmates on any group project.

16. Classroom Morale. Classrooms are not usually random samples of the whole school population, but have distinctive characteristics in their respective tendencies to be of service. This code of the classroom seems to accumulate from year to year, moving either toward greater or less coöperation according to the majority influence.

17. Sunday-School Attendance. Those who have at any time been enrolled in Sunday school are more coöperative than those who have not been, but the difference is very slight. Those regular in attendance are more coöperative than those who attend only occasionally. There is a small correlation (.10) between length of attendance and service, but against this fact must be set the peculiar tendency in population X for those who had been in Sunday school for eight years or more to be less coöperative than the rest. Whether the Sunday school is to be held responsible for either result is not determined.

18. Club Membership. In the case of one boys' club and one girls' club, there was a slight tendency for members to be more coöperative than non-members. In the case of the boys' club, there was a corresponding tendency for the members to reflect a

higher cultural level. This tendency for club members to excel was not present on every service test, but was, instead, reversed on several. In the case of a religious society for young people in population Y, the superiority of members over non-members was somewhat greater than that found in the other clubs.

19. Motion Picture Attendance. Considerable difference between the average service scores of regular and occasional attendants was noted. In the Z population the distinction lay between those who went less than once a week and the rest. In Y it was between those attending not more than once a week and the rest. In X the line of division was between those attending not more than twice a week and the rest. In Z it was the small non-going minority who were the most coöperative. In X it was the small minority attending three times a week or more who were the least coöperative. In Y the numbers were more equally divided. Whatever may be the effect, if any, of attendance, it must vary greatly with the population involved. Presumably the differences are due to some selective factor not isolated by our technique.

20. Suggestibility. The case for the power to resist suggestion is about the same as for intelligence (an r of zero in X and Z and .25 in Y). At the extremes of the suggestibility scale the difference in service tendencies is quite marked in population Y, the more suggestible proving far less coöperative and the less suggestible far more coöperative than the average. This condition was not found in X or Z.

21. Cleanliness. As with similar facts, the association with service is positive but slight. There is a concomitant tendency for the very clean to be also considerably more intelligent than the others, as well as somewhat more coöperative.

22. Sociability. Contrary to other facts so far treated, the sociable tend to be in the middle of the service scale rather than at the top.

23. Friendship. Pupils tend to resemble their friends considerably more than they resemble their classmates in general, particularly if the feeling of friendship is mutual and the friends are found in the same classroom.

24. The Effect of Majorities. The majority sex in a classroom tends to be more coöperative, just as it also tends to be more deceptive. On the other hand, the majority or predominant national and religious groups tend to set the pace for the rest in either direction.

25. The Effect of Teams. In one group the two teams chosen on the occasion of a test to compete with one another induced far more unselfish loyalty than did the class as a whole when pitted against other classes in the same school. The vividness of group consciousness brought about by this spontaneous and visible organization may be of considerable importance for the control of conduct.

26. The Effect of Repetition. In one group, when the chance to give one's work to the class was repeated the following day, the class made far less appeal as an object of unselfish loyalty.

27. The Instability of the Coöperative Motive. In the course of the extended efficiency test, the motive of class loyalty decreases more rapidly than the self motive. Also, the behavior of individuals working for others or when fatigued is far more oscillating than when working for self or when fresh. We also found in one population that children who voted to keep a money gift for themselves were more apt to stick to their first judgment than those who voted to give their money away. The stronger motive is the steadier.

28. Service Not a Unified Trait. As in the case of deception, no evidence of the existence of a general trait of selfishness or unselfishness was found. Behavior in different situations calling for coöperative or charitable behavior was only loosely organized.

29. The Effect of Teaching. In one experiment conducted on two groups, a carefully planned series of stories and discussions leading to a definite formulation of policy by the pupils was followed by a significant increase in scores on the Efficiency Coöperation and Free Choice tests on the day following the three days on which the teaching occurred. The superiority of the experimental over the control groups was comparable to that of one class in population P which had been studying the story of Heidi just before

our tests were given. The pupils who had been with this teacher the year before, however, although she had used this story with them, proved no more coöperative than the rest in their new classroom. The problem of the permanence of the effect of such teaching is thus not solved.

30. Amount of Unselfishness. In activities involving a conflict between self and others, as in the Kits and Free Choice tests, the percentages of children in the three populations who did *not* share at all with the others are, for the Kits, 38, 36, and 60, in X, Y, and Z, and 44 in the entire population, and for the Free Choice, 32, 27, and 24, in X, Y, and Z, and 27 for all. When actually sharing with others, the amount given away depends, obviously, on the strength of the appeal and the attractiveness of the objects to be given. In the case of the Kits about two objects in ten were given away, and in the case of the Free Choice about half the units of work were given away. When working alternately for self and class, 35% of the children did more for class than for self. On the whole, most children may be regarded as either moderately selfish or moderately unselfish according to the standard by which they are compared. Certainly they cannot properly be regarded as universally selfish or the reverse.

Summary. The thirty statements just made follow the order in which the forty variables referred to were treated in the text. A classification of this material will assist in gathering its implications in one comprehensive conclusion.

We are dealing with two types of data. The first concerns individual differences and their causes. The results of our study of service with regard to this type of data are expressed in correlations and in the comparison of extremes. The second concerns group differences and the causes which operate to effect changes in groups. The results of our study with regard to these facts are expressed in terms of differences between means or medians and correlations of group means or medians.

Individual Differences. With regard to the association of individual differences with such facts as permit considerable individual variation, we find that the most important factor in determining

the service behavior is the mutual friendship of children in the same classroom. Next in significance is satisfactory school adjustment as represented in adequate intelligence in comparison with one's classmates, suitable grading, and such relationship with the teacher and one's classmates as finds expression in academic and deportment marks and classroom morale. Third is the influence of the home, whether biological or social, as shown by the resemblance of siblings, the effect of parental example, and such cultural factors as are bound up with occupational level, treatment of children, knowledge of etiquette and general cultural information, and good habits in the care of one's body. Finally, there are a few miscellaneous facts only loosely related to service tendencies, such as sex, physical and emotional condition, power to resist suggestion, and sociability.

Group Differences. With regard to the general effects of experience, or the effects of such selection of cases as may be associated with certain types of experience, we find that the largest group differences are between communities, religious groups, and national groups and between those who for whatever reason attend motion pictures less or more than the bulk of the population of which they are a part. There is a slight tendency for Sunday-school and club members to be more coöperative than non-members, but day schooling does not have any consistent effect. Being in this classroom or that is important in determining behavior, but being a member of a team is still more significant. The general level of coöperative behavior can be raised at least temporarily by suitable teaching. On the whole, the outstanding facts are facts of population selection and structure, the better established community with greater homogeneity in race, religion, and intellectual ability proving the most coöperative in spite of all other differences which might tend to make it less coöperative.

Character of Service. Finally, regarding the nature of service tendencies, children apparently can be arranged on a scale of behavior just as on a scale of height or weight, with the majority being at or near an average of helpfulness and a few at each end being either very selfish or very unselfish. Class loyalty is usually

as strong a motive as the desire for reward or recognition, and generosity is about as prevalent as greed. The tendency to be of help is oscillating, however, particularly with the middle 50% of the cases; it also tends to fall off under fatigue or upon repetition of the situation demanding assistance.

IMPLICATIONS

As in the case of tendencies to deceive, it is apparent that such efforts to train children in forms of charitable and coöperative behavior as are ordinarily used in school have very little, if any, effect. Even when presumably specialized efforts are brought to bear on the problem, as in Sunday schools and clubs, the effects, if they be effects, are not great enough to make members stand out as being consistently superior to non-members. The fact of sex, of the kind of friends one has, of the occupational status of one's parents and the recency of their arrival in the United States, as well as the country from which they came and the church with which they are affiliated, of time spent at motion pictures, of success or failure in school, and of emotional stability or instability — all such facts seem to have more to do with existing individual differences in helpful conduct than do age or schooling or special forms of training.

On the other hand, it seems to be somewhat easier to change conduct of this character than to change deceptive behavior. There are several possible reasons for this. One is that deception is a positive, active response, and any change for the better means doing *less* of it, not more. The momentum is already established, and what is required is to stop moving. In the case of service, which is also a positive, active response, any change for the better means going on in the same direction, or at least starting. The honest person does *nothing* to deceive. The unselfish person must do *something* to help.

We found that in the matter of deception children were sensitive to slight changes in the situation, but that it was not easy to keep

from deception those whose dishonest tendency was well established. We also found that changes in the situation, as well as discussion of the relations of team mates to one another, resulted in marked increases in coöperative activities, at least temporarily. Upon the repetition of the same situation, however, the unselfish tendencies fell off.

A second reason for the relative ease with which coöperative tendencies seem to be affected may lie in the surreptitious nature of deceit, which makes any change a more fundamental and individual problem. A child who becomes more honest may feel that it is at the expense of his standing in class and that, since his change in conduct is not known and receives no social approval or support, the game is not worth the candle. Those who are more coöperative, however, as in the case of the coöperation experiment reported, are working for a stake to be won, not lost, through the improved or more unselfish behavior. In this case, however, the advance is probably only from one form of self-interest to another. In the case of the Envelopes and Kits, there is at least the satisfaction of bringing joy to others, even if at some expense to oneself.

In planning to improve the coöperative and charitable behavior of children or in setting up a definite experiment which might have such improvement as its objective, it is obvious that one of the most important phases of the problem will be the uncovering of the existing concomitants of such social service tendencies as are already present. In some cases the children may be somewhat surfeited with social service activities. In others such enterprises may be a genuine novelty. In some cases service may involve sacrifice. In others it may be a mere pastime. In some instances the whole cultural background may prove to be an inhibiting influence. In others it may powerfully reinforce any suggestion for coöperation. In some schools the tradition may be toward selfishness. In others it may be toward unselfishness. The problem is not simple, but with due consideration given to forces and motives, as well as to the particular situation in relation to which unselfishness is to be achieved, there is no doubt but that changes in social conduct can be produced.

Whether such changes are of significance for character is another question the consideration of which must be postponed.

Such needed light on the nature and causes of individual differences in tendencies to be of service as the facts reported in this book have failed to throw must needs be derived from the extension of research into this vast field by means of many other scientifically controlled experiments.

BOOK TWO

Studies in Self-Control

PART I

The Measurement of Self-Control

CHAPTER XVI

ANALYSIS OF SELF-CONTROL

INTERMINGLING OF PERSISTENCE AND INHIBITION

In our Introduction we mentioned the intermingling of such personality characteristics as helpfulness, persistence, and inhibition in any concrete response. If it were possible to isolate the factors in a situation which are associated exclusively with a single type of response, or to control a series of complex situations in such a way as to vary only the factors associated with a single type of response, our problem would be much simplified. We have made only rough approximations to this standard. But even if one cannot separate the elements of behavior in actual life, one can at least talk about these elements and attempt to gain a sufficiently clear idea of them to serve as a guide in further experimentation.

In this book we are dealing with the two types of response, which we have called "persistence" and "inhibition," including them under the general term of "self-control." We are not concerned to defend our terminology, but only to explain it. Our assumption is that, whether one persists in an act or refrains from making a suggested response, his action can be explained in terms of his previous experience unless he is under some form of external compulsion, which is very rare and almost impossible to administer. Violence is usually exercised through the medium of threats — that is,

through the compelling power of experience which gives meaning to the threats. This is true whether one is restrained from action or compelled to action. This impingement of past experience upon present experience is the substance of what we popularly call self-control. It is the accumulated heritage of what has happened operating to determine what shall happen next. This stock of controlling experience is peculiarly individual. It is never exactly duplicated. It is what gives character to the self. It is a part of the self.

THE FUNCTION OF INTEREST

From another point of view we may think of both persistence and inhibition as functions of interest. If some one term is needed for the dynamic human factors in any situation, " interest " is as good as any. It is hardly an adequate explanation of an activity to call it an interest or to refer it to some interest as its cause, but no empirical treatment of behavior would be possible without some consideration of the sources of energy which are utilized in any activity. Whether we agree with Woodworth's claim that every habit system is also a drive, or with McDougall in tracing all energy and motive to the fundamental instinctive organization of the human body, we must recognize in every act a tendency toward some objective, which, unless thwarted, is carried through to completion.

In our *Studies in Deceit* we proposed the theory that drives toward objectives are functions of the situations which excite them. No matter what may be the individual's potential capacity for action, the urge to act is itself a specific adaptation which is characteristic of the individual in the specific situation involved. We reported how a child would tend to overcome a certain amount of resistance in order to cheat on a test. If a series of resistances is presented, his drive will carry him through them all up to the point where the resistance is too great for his tendency to master it but will not succeed with any greater obstruction.

In the case of intelligence, the greater resistances offered by tests of increasing difficulty represent levels of capacity which

no change in motivation can affect. No conceivable reward would enable an inferior mind to grasp the theory of relativity. But in the case of our deception tests, the difference among the subjects was not one of capacity to deceive. In this respect they were all equal, as all that was involved was the changing of a pencil mark. The difference was one of interest or, better, of a balance of interests. Some wanted to cheat more, or had more to gain by cheating, or had less to lose by cheating, or cared less about honor than about marks, or feared punishment for low grades more than the pangs of conscience.

This statement possibly oversimplifies the situation, as it is possible that some minds are incapable of great enthusiasm for standards of honor, or are less able to comprehend them, than are other minds. That is, capacity of one sort or another may be involved. But it is involved through the medium of interest and not directly in the act itself. The individual differences in the actual situation are differences in amounts of interests.

By this illustration from the practice of deception, we can see in what sense persistence and inhibition are both functions of interest. Absence of interest is absence of activity. Increased interest implies more prolonged activity. The interest may be in the satisfying quality of the activity itself apart from its more remote results, as when a child keeps on swinging, or may be in the end to be gained, even against the resistance of a disagreeable or painful activity required to achieve the desired end, as when a woman allows a dentist to torture her for an hour in order to have her teeth made more attractive. The first illustration excludes almost entirely the factor of inhibition, although undoubtedly a youngster will sometimes need to inhibit the tendency to please his mother or to eat his dinner in order to prolong the pleasure of swinging. In the second case, persistence and inhibition are so closely interrelated that they constitute in a sense the two tendencies which, in merging, issue in the behavior exhibited. There is a nice balance between the desire for the end set up and the control of the desire to escape the necessary pain. Now and then the pain conquers and the work ceases while the patient re-

covers her interest in the object to be gained. Her persistence in the dentist's chair depends on what is at stake, on the amount of interest in the improved teeth.

Other complicating factors are such matters as reputation for being a good sport, a personal ideal of stoic indifference to pain, and, in such cases as that of dental work for health rather than beauty, a kind of grit or determination to do whatever duty may require. But whether one refers the willingness to keep on with an act to pleasure in the act or to desire for its object, and whether the act be inherently agreeable or disagreeable, the factors involved are interests, or attachments to a great variety of associated consequences, real or imagined, which are of moment to the self.

We may think of any response, then, as representing the activity of certain tendencies and the inhibition of other tendencies. In some cases the tendencies incited to action are approved by the individual, and in some they are disapproved. In the first case, there will be conflicting tendencies which, if allowed to proceed, would interfere with the approved tendencies. In the second case, the tendencies directly incited are those which must be inhibited by others if the individual is not to do what he disapproves. The motives, in either case, may be varied. A tendency may be approved or disapproved for many reasons which do not appear in our test results.

VARIETIES OF INTEREST USED IN THE TESTS

It may seem from this discussion that we cannot deal adequately with either deceit or service without taking into consideration those interests which make for persistence in the chosen activity and for the inhibition of conflicting tendencies. We have already shown in our presentation of the results of the tests how those who are willing to give away scores at personal sacrifice for the sake of the class are usually, but by no means always, the persons who will do the most work for the class when continuous effort is demanded in order to be of service. This corresponds to the difference we often note among persons of charitable inclinations. Some will

give money which they would otherwise spend on themselves, but will not exert themselves in any personal way. On the other hand, some will work hard for a cause, but only when no personal sacrifice is involved. Whether this difference is related only to the desire to be of service or is displayed also in situations involving no coöperation at all is an important problem.

SELF-CONTROL AND CHARACTER

Traditionally, self-control has been included among the virtues of the civilized man from the beginning of history. The writings of wise men, whether of the ancient cultures of China and India, of the Nile valley and Asia Minor, or of Greece, exalt industriousness and temperance, in evident fear lest the indolence and self-indulgence of the ordinary man prove his undoing. So successful, apparently, were such half legendary figures as Confucius, Gautama, Moses, and Mohammed, the forerunners of Puritan ascetics, in enthroning these virtues of an economic social order, that the natural desire of human beings for the comforts and excitements of life became actual sins in the eyes of the people, needing to be excused and justified by the rationalizing arguments of the Omar Khayyáms of each generation.

In making a virtue out of self-control, however, the Puritan conscience-monger was as much self-deceived as those who build philosophies with which to clothe their natural desires. For the chief difference, after all, between the civilized and the uncivilized lies not in the possession of a general virtue of persistence or a general power of inhibition, but in the development of certain habits rather than others. The savage, at least according to the reports of those who have studied his customs, is able to perform prodigious feats of endurance, whether of physical strain or of pain; he has acquired self-control in certain directions quite beyond the power of the civilized man who preaches to him of temperance and industry. In other respects, however, as in the accumulation of goods, or in the use of alcohol, he is less disciplined than those who make a virtue of property and sobriety. Yet the motive for the

possession of goods may be far more selfish and the results of their accumulation may be far more destructive to the peace and growth of the individual than the more simple attitudes of the untutored.

Some recent studies have tended to revive the old interest in inhibition as the very core of character. Roback,* for example, devotes considerable space in his admirable survey of character study to the doctrine of inhibition. His definition of character is "an enduring psycho-physical disposition to inhibit instinctive impulses in accordance with a regulative principle." Our own view is that whether one emphasizes the presence of the regulative principle or its inhibitory effect depends on the interest of the moment. Both the factor of pursuit of ends or adherence to principles or sensitiveness to continuing motives is present as well as the capacity of any of these factors, or of the individual as a whole, to inhibit contradictory tendencies.

These brief comments will serve to remind the reader once more that we do not regard tests of specific behavior tendencies, such as persistence and inhibition, as measures of character. We shall try to indicate in a subsequent volume the extent to which the presence of various degrees of persistence and inhibition in certain situations correlates with character as a whole.

---

* Roback, A. A., *The Psychology of Character*, Harcourt, Brace and Company, 1927.

CHAPTER XVII

PREVIOUS ATTEMPTS TO MEASURE INHIBITION AND PERSISTENCE

During the past twenty years both psychologists and physiologists have taken considerable interest in the subject of inhibition and persistence. No fewer than three hundred books and articles have appeared which deal with various aspects or phases of these tendencies. Most of these studies have been conducted by physiologists whose main interest is in the nature of the inhibitive process and the conditions under which it occurs.* Our present study, on the other hand, concerns the measurement of individual differences in inhibition of the type popularly known as self-restraint. In this chapter we shall briefly review the tests which have this object in view. These may be conveniently classified according to the kinds of responses inhibited.

CLASSIFICATION OF TESTS

Measures of the Ability to Inhibit Emotional Tendencies Such as Fear and Anger. Crane † attempted to measure the inhibition of fear responses by determining how many times in a certain number of trials the subject would keep his hand in an apparently dangerous place, even when told that injury was impossible. The apparatus used resembled a pile driver (Crane called it the guillotine). It consisted of two upright standards with an apparently

* For a summary of such material, see, *e.g.*, Dodge, R., "The Problem of Inhibition," *Psychological Review*, 1926, **33**, pp. 2 ff.

† Crane, A. L., "Race Differences in Inhibition," *Archives of Psychology*, No. 63, 1923.

heavy weight mounted to slide up and down between them. The weight was in fact very light, and the standards had a safety device in the slots so that the weight when it fell would come to a dead stop about one inch above the hand. But to the subject the apparatus looked very dangerous. After being assured that he would not be injured in any way, the subject was requested to place his hand on a board directly under the weight. The test was whether he would withdraw his hand when the weight started to fall.

Crane recorded the number of trials necessary before the subject would leave his hand in the apparatus, and the number thereafter before he would cease to flinch when the weight came down. His object was to study race differences by testing 100 whites and 100 negroes. His results show that a larger percentage of the whites (72%) accepted his statement that they would receive no injury, and let the hand remain on the very first trial and all subsequent trials, against 62% of the negroes; yet, on the contrary, the negro subjects ceased to flinch on the average in fewer trials than the whites.

His tables show interesting individual differences within both racial groups in the number of trials required to overcome the tendency to flinch.

Moore * endeavored to measure the extent to which persons doing mental multiplication would be distracted by certain emotions. After preliminary work at mental multiplication had eliminated practice effect and norms had been established for working without distractions, the following emotional distractions were applied:

Anger distractions. (1) The subject was given a letter containing an unjust accusation just before the multiplying began. (2) During one problem an assistant slapped the face of the subject and flipped his nose. (3) An assistant tried to interfere with his mental work by saying aloud a series of numbers.

Fear distractions. (1) Reference was made to snakes; then a

* Moore, H. T., "Laboratory Tests of Anger, Fear and Sex Interest," *American Journal of Psychology*, 1917, **28**, 390–395.

rubber tube was wound around the subject's neck. (2) A concealed assistant placed an icy hand on his forehead. (3) A falling chair distraction. (4) An expected electric shock.

Sex interest. The subject was asked to do his multiplying in the presence of nude pictures.

Repulsion. Objectionable odors and objects repulsive to look at or to feel were presented to the subject while he was multiplying.

Embarrassment. The subject was required to multiply while facing a roomful of watching students.

Score. The average time for twenty problems was taken. All distractions were combined. This was a general distraction time. Then the anger, fear, sex, repulsion, and embarrassment distraction times were computed separately and expressed as percentages of the general distraction time. Twenty-two subjects ranked themselves with respect to these emotions. The correlations between self-ranking and test-ranking were .37 to .51.

Fear had the highest average distraction time, then came anger, embarrassment, sex, and repulsion. The r between the distraction times of fear and anger was $-.48$. All other intercorrelations were low.

Measures of the Ability to Inhibit an Interest. In the study already referred to, Crane sought to measure the ability to inhibit the curiosity impulse by using a distraction technique. He first had his subject count symbols written on plain cards and later count them when written on cards also bearing interesting pictures. The difference between the scores represents the ability to inhibit, the larger the difference the less the inhibition.

Burtt * devised four tests for measuring interests objectively, one of which is similar to Crane's test. He used two passages of prose, one dry and the other presumably interesting. Each passage had irrelevant three-lettered words scattered throughout. The task was to cross out the irrelevant words as rapidly as possible. The real task was whether or not the subject could inhibit the tendency to read the interesting passage.

* Burtt, H. E., "Measuring Interest Objectively," *School and Society*, 1923, **17**, 444–448.

Measures of the Ability to Inhibit a Fixed Habit. The battery of the Downey * Will-Temperament tests contains one test called "Motor Inhibition." It consists in writing the phrase "The United States of America" as slowly as possible. Dr. Downey says, "The object of this test of motor inhibition is to determine the examinee's power to hold back a motor impulse." Uhrbrock † made a careful study of the reliabilities of the Downey tests and reports a self-correlation of .66 for this one. This would indicate that this sort of test has possibilities as a measure of one type of inhibition.

Measures of the Ability to Overcome an Obstacle or Inhibit an Annoyance. The Downey battery also contains one test of this variety. In this test the subject is asked to write with the eyes closed. The examiner then places a block (a fountain pen or a small object) in front of the pen or pencil of the subject, pressing down firmly so as to form a real obstacle. Since this test could not be adapted to group testing, it is omitted from the battery of group tests; hence little information is available as to its reliability.

Moore and Gilliland ‡ devised a series of tests to measure aggressiveness. Four of their five tests are based on the hypothesis that the more aggressive subjects will succeed better in overcoming annoyances than the less aggressive. Their scheme was to have the subject start some kind of standard mental work, such as mental addition, and then introduce various annoyances and distractions and measure their effects by the difference in score on the arithmetic work when done with and without the annoyances. The annoyances used were: (1) staring the subject in the eyes, (2) an electric shock, (3) a dead snake. The examiners found wide individual differences in abilities to deal with these annoyances, and that those rated as the most aggressive were the best inhibitors.

---

* Downey, June E., *The Will-Temperament and Its Testing*, World Book Co., 1923.

† Uhrbrock, Richard S., "An Analysis of the Downey Will-Temperament Tests," *Teachers College Contributions to Education*, No. 296, 1928.

‡ Moore, H. T., and Gilliland, A. R., "The Measurement of Aggressiveness," *Journal of Applied Psychology*, 1921, **5**, 97–118.

In a series of tests devised to measure the introversion-extraversion qualities of young children, Marston* has one called "Social Resistance." This is a standardized situation in which the child has certain resistances to overcome in getting acquainted with the examiner. It is really a test of the ability of the child to overcome social barriers.

Morgan and Hull† report a technique similar to the Downey Block Test, in which they use a pencil maze to be traced through with a stylus. The work is made difficult by inserting blocks in the various alleys. By using a series of blocks the maze may be rendered insoluble. The test was not scored quantitatively, but the behavior of the subject was observed and classified into one of nine groups or categories graded according to the degree of persistence shown.

Measures of the Ability to Inhibit the Tendency to Quit a Difficult or Disagreeable Task. The tests just listed all assume that the activity engaged in is satisfying. Some annoyance is injected the effects of which are to be inhibited. With these tests the opposite is the case: The activity engaged in is disagreeable and some interesting or satisfying distraction is introduced. The task is not to quit. Such is Marston's compliance test, of the same series as the Social Resistance test already mentioned. In this the child is told that there are some toy ducks in a box or case and that, if he will open it and get them out, he may have a game with them. The case containing the ducks is fastened in a complicated way and is very difficult to open. The question is: Will the child attack the problem and stick to it without quitting or asking for help?

Gilliland‡ used as one of the indications of aggressiveness a staring test, in which the score is simply the number of eye movements observed by the examiner. In such behavior as trying to

---

* Marston, L. R., "The Emotions of Young Children," *University of Iowa Studies in Child Welfare*, Vol. III, No. 3, 1925.

† Morgan, J. J. B., and Hull, H. L., "The Measurement of Persistence," *Journal of Applied Psychology*, 1926, **10**, 180–187.

‡ Gilliland, A. R., "A Revision and Some Results with the Moore-Gilliland Aggressiveness Test," *Journal of Applied Psychology*, 1926, **10**, 143–150.

"outstare" the examiner or any other person there is in many individuals a strong tendency to look away. Hence success at such a task involves clearly the ability to inhibit this tendency.

One of the few inhibition tests that is now available for use is Fernald's Achievement Capacity test.* The subject is requested to stand on his toes with his heels off the floor. The score in the test is the length of time a person can do this. There is a special apparatus consisting of a pivoted plate connected by a wooden lever to an upright arm, with a needle arrangement at the top to measure the oscillation and a bell attachment so that, if the subject lets his heels down below a certain point, the bell will ring.

The test is given by letting the subject step on the plate and stand on his toes as long as he can. Fernald found that a group of reformatory prisoners averaged 17 minutes, 18 seconds, whereas a normal group of comparable ages, weights, sizes, etc. averaged 52 minutes, 30 seconds.

In discussing this experiment, Fernald says: "Experience in applying the test to 128 subjects shows that in every case the decision to give in preceded the necessity of it. One subject freely admitted he could have stood longer but said in reply to a question, 'There is nothing in it for me.'" This is taken to mean that he did not have enough mental persistence to try. Fernald thinks that this test affords a means of measuring those mental qualities or forces which make for success in competition. (Incidentally, it might be questioned whether such apparently meaningless persistence has any moral significance — which again may throw light on the nature of competition.)

A test that resembles our persistence technique is one reported by Chapman.† It is a group test and adapted for school use. The task set is that of word building and is adapted from the word-

* Fernald, G. G., "An Achievement Capacity Test," *Journal of Educational Psychology*, 1912, **3**, 331–336. The apparatus is supplied by the C. H. Stoelting Co., Chicago.

† Chapman, J. C., "Persistence, Success, and Speed in a Mental Task," *Pedagogical Seminary*, 1917, **31**, 276–284.

building test reported in Whipple's *Manual.** Chapman secured three scores with this test: first, the total time spent before giving up — which he called the " persistence " score; second, the average time per word to build the first ten words — calling this " speed of thought "; and third, the number of words completed — or " success." The maximum number of words possible with the six letters used was 54, of which 33 were within the vocabulary of eighth-grade pupils. The three eighth-grade groups tested, however, scored only an average of 14.6 words, and their persistence time ranged from less than one minute to twenty minutes, with a median at ten minutes. He found that the more rapid workers show less persistence but more success. The average correlations for three groups are: speed and success, .28; speed and persistence, −.29; persistence and success, .65.

* Whipple, G. M., *Manual of Mental and Physical Tests*, Warwick and York, 1910.

CHAPTER XVIII

CLASSROOM TESTS FOR MEASURING PERSISTENCE AND INHIBITION

PERSISTENCE

On the assumption that persistence is in part a function of the subject's interest in the activity which is being pursued, we planned a series of tests representing a variety of interests. The first of these was curiosity as to how an incomplete story really ended. The second was the solving of a mechanical puzzle. The third was the solving of a mental puzzle. We shall describe these in turn as they developed.

The Story Resistance, or S–R, Tests. Three stories are used, two of escape from danger and one of adventure. The first story is read to the class up to the climax, and the ending is supplied written in such a way that it is very difficult to read — the difficulty increasing as the end is approached. The child is given the option of finishing the story or beginning a new one which ends in the same difficult way. Again he is given the option of finding how this story ends or beginning a new one.

Three degrees of reading difficulty are provided in the ending to each story. The first level consists of capital letters run together with no spaces between words, as this: CHARLESLIFTED LUCILLETOHISBACKANDTOLDHERTOHOLD. A second order of difficulty is secured by mixing small and capital letters, as this: ChaRlEslIFtEdlUCiLleTohISbACkaNDToLdHeR. A third order of difficulty is reached by using both small and capital letters with spaces in the words like this: Char LEs LiF Ted LuCIlLEtO HisBacKa NDtOlDhErTo HoLd. A short practice period is

given in reading this kind of material before the first story is read. The aim of this is to accustom the child to reading such material and to marking off the words and also to get a measure of his ability in doing this kind of thing.

The instructions (Appendix E) are to place a vertical line between the letters to show where one word ends and the next one begins. This makes reading somewhat easier and also enables the examiner to determine how far each child has read. The following quotation illustrates the final form of this material.

Charles and the Terrible Moment

In administering the test, the story was told up to a point of chief suspense. It is about a boy and his small sister who were crossing a trestle when a train was heard approaching. The boy had a bruised wrist.

"Finally he could see the trestle, and there in the middle of it, walking slowly, heedlessly, along, was Lucille! He started to run towards her. At that instant the train whistle sounded again, this time much nearer. Lucille turned at the sound. She realized her danger. A queer helpless panic set in. 'Charlie,' she cried piteously, 'I want you, Charlie!'

"Charles reached her at last, just as the train appeared in sight around the bend, not two hundred yards away. There was no time to walk, or run, the rest of the way. The trestle was not wide enough for him to stand beside the tracks and hold Lucille, and far below was the shining cruel depth of the river.

"They seemed trapped. He could hold on for himself, but what about Lucille? Would his bruised wrist bear his own weight, let alone Lucille's? If he only had a rope, or if Lucille had a belt or strap, by which he could lash her to a tie!

"Again the terrible piercing shriek of the whistle screamed at them. Charles could see the frightened face of the engineer . . ."

At this point the examiner says: "How many would like to know how this story ends? I am going to let each of you find out for yourself. The sheets I shall pass out have the ending printed on them, but the words are run together just the way they were in the exercise you did a moment ago. You can best read the story by drawing a short line between each two words just as you did before. . . ."

A sheet was passed containing the rest of the story printed as follows:

Copyright 1927, Institute of Educational Research, Teachers College, Columbia University.

Form S-R-2

Put your name here

CHARLESLIFTEDLUCILLETOHISBACK"PUTYOURARMSTIGHTAROUNDMYNECKANDHOLDONHOLDON
DOYOUHEAR?"THENHESWUNGHIMSELFOVERAPROJECTINGTIEANDWITHGREATDIFFICULTYLETHIMSELF
HANGCRASHTHEENGINEROARED PASTLUCILLELETOUTACRYOFINTENSEFEARCHARLESTRIEDTOCALL
TOHERTOREASSUREHERBUTHISVOICEWASDROWNEDINTHEROAROFTHEWHEELS.

PERILOUSLYDANGLINGOVERTHERIVERTHETWOREMAINEDWHILETHELONGTRAINPASSEDCHARLES
GRITTEDHISTEETHHEDIDNOTBELIEVEHEWOULDBEABLETOSTANDTHEPAININHISHURTWRISTANDTHE
STRAINOFSUPPORTINGHISOWNWEIGHTANDLUCILLE'STOOTHELITTLEGIRL'SSTRUGGLESONLYMADE
THESITUATIONWORSEANDSHECOULDNOTHEARHIMWHENHEBEGGEDHERTOSTOP.ATLASTIT
APPEAREDTOHIMTHATHEMIGHTFAINT.

HECLOSEDhISEYesWHenHeOPenEDTHemHEDIScOvEreDTHaTTHetRainHADgOnE.

NoWhoWTogETBaCkONthETRESt1e.HoWTOBRingTHaTTErrIfIEDBURDeNOFAChILDuPtOSafeTY
WHeNhEhIMsELFwASaLMostexHauSTEDHEtRIEDtOlIFThIMsELFtHOUGHhEwASVErYsTRONGfORabOYoF
HIsAGEhEfOuNdNOWTHATheCOULDnOtraisEhiMSELFandlUCILLEtOOhECOULDoNLYcLInghEcOmmENCED
TOsHOUTnOWLAYingHISHEADwEARilyAGaINSTHiSARMHeCALledFORhELPhECOULDhEARHisLIttLE
SIsTerSObbINgANDTHeSOUNDgAVEHIMsTReNGtHTOrENeWHiSOwNsHOUTSwOULDNooNeeVeRCoMe
WerETHeYToDIeaFTEraLL?

fiN ALly tAp-tAPC AME ARHYTH Month eBriD GeruNNing fee Tfee Tcom INGtow
ArdT Hem Cha Rles TRi Ed oN Cem ORE TOS Hout ButheH adgRow Ntoo We AkiT WAsaT
Rain MaN WHOHA DRun BackfrOMT HE StoPp EdT Rain ToiN VESTig AteH is Mat URe
StR oN GH And sQuicKLYR Each Edd own An DGoTH OLD Of the Han Gin Gfo RmssLoWly
THEy We Red Raw Nup Ward. CAR ef Ully THe y weReL aid on The Cross TiESoFt He
TRe Stlef or Amo Me NtN Eitherof The Mo OuLdsp Eak. THeHusKYT Rain man Loo
KED Ad Mi Ring LyAt oHaR les. We llNo Wit's Ab Ravel Ad Year e He Com Men Ted
I'llbe TTH IS Her EKid' sMother WillBe A'THanK in' You ALL Rightco Me Al on
gWiThmE An DI'LL He Lpy oua Cro SSN OW.

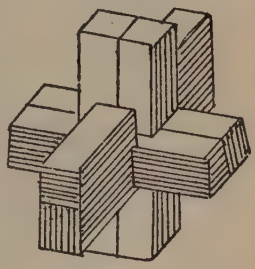
Scoring the Story Resistance Tests. A uniform time of three minutes was allowed on the practice exercise. The score on this was the number of words of text spaced off in the way just described. When the end of the story was presented, the length of

time each child worked on it was noted at the time, and later the number of words he spaced off was counted. The score was the number of words spaced off in the story divided by the number of words spaced off in the practice period of three minutes. It is an approximate measure of the time (in three-minute units) the pupil really worked on the story.

Although the actual time spent on the story was not needed for this score, it was kept as possibly offering a second type of score.

The Puzzle Mastery Tests. As we have already suggested, two types of puzzle were used, mechanical and mental. In these tests we tried to arouse the desire to achieve or master, so far as this was associated with puzzles of this sort. After much experimentation we settled on two types of material, the Cross and Ring tests and the P–M, or Puzzle Mastery, tests.

The Cross and Ring Tests. In population C, where we had been doing considerable testing, we tried out our two mechanical puzzle tests. The Japanese Cross was a mechanical task requiring that six irregular pieces of wood be assembled to make a three-dimensional cross. The accompanying figure is a drawing of the assembled puzzle. It is well known to be exceedingly difficult, and we hoped that none of our subjects would solve it in the time allowed. In 850 cases ranging from grades 5 to 8, only 2 children mastered it.



When we first administered this test, we passed the puzzles out and told the pupils to work at them until they felt like stopping, when they were to raise their hand and sit quietly. At five-minute intervals the examiner reminded the pupils that they might stop whenever they felt like it and spend their time on their regular school work. As the examiner collected the cross, he noted the time spent and told the pupil to go to his own work.

We found that this indefinite "pull" offered by the child's regular work introduced an element of uncertainty into the test; and furthermore, the task itself was narrow in its demands upon

the subject. Subsequently, therefore, we introduced a second puzzle, called the Chinese Rings puzzle. This had to be taken to pieces before it was put together and so offered a contrast with the wooden cross, which was presented in a " knocked down " condition. Each child was given his choice of the Cross or Ring and was allowed to stop working on whichever he chose first and go to the other whenever he pleased.

His persistence score was the time spent on the first puzzle. We first allowed 30 minutes for this test but later extended the time to 80 minutes. Only 8 of 850 subjects got the Ring puzzle apart, and none put it together again.

We also provided what we called " fillers " for those who got tired of both puzzles before the rest were through. These consisted of paper-and-pencil puzzles of various kinds. Thus each subject had the same type of activity as every other both during the test and as an alternative to the test.

The P–M Tests. Our first series of paper-and-pencil puzzle tests* contained the following problems:

1. The Traveler's Problem. A conventional map consisting of three concentric circles, joined by several radial lines, on which twenty " towns " are located. The problem is to start from one town and visit all the others without going over the same road twice.

2. The Broken Chain Puzzle. A chain of fifteen links was pictured, the links being fastened together in five groups of three each. The subject was to find how the groups could all be fastened together by opening only three links.

3. The River Puzzle, similar to the fox, grain, and man puzzle. " Four boys arrived at a river and found a boat that would carry just 200 pounds but no more. None of them knew how to swim; so they had to manage to cross in a boat. Tom weighed 200; Jim weighed 200; Bill weighed 100; and Dick weighed 100. How did they get across? "

---

* Most of these may be found in *Modern Puzzles*, a collection by H. E. Dudeney, published by Frederick A. Stokes Company.

4. The Magic Square Puzzle. The example below was given. The subject was instructed to make a similar square, with four cells in each direction instead of three, which would add up to the same amount in all directions, with a different number in each box.

Here is a Magic Square:

8	1	6	=15
3	5	7	=15
4	9	2	=15
15	15	15	15 15

Thirty minutes in all was allowed. As each child completed or tired of one he brought it to the desk to receive the next, and the time was noted on his papers so that the amount of time he worked on each, with or without solving it, could be computed.

When a pupil handed in the Magic Square test, he was given a routine task consisting of counting the number of letters in a page of lines of miscellaneous letters like this:

wesaqwert dcfg vgbhj yuiom awerty xdcfvg _____
kopy nj vg cf sawe erty liko bh ftgui dswe vgh uiop un o fr fr _____
qwertyunioopfvds gv gyui j uiop vgne sedft scvrd crop desertyi _____

We found, again, an ambiguity in the results, owing to the fact that some solved the problem and some did not. This greatly varied the interest in the task. So we adopted in our final program only the very difficult Magic Square and a Magic Word puzzle. These, which we called P–M 1 and P–M 2, respectively, are shown in Figures 67 and 68.

Each child was given first P–M 1, and when he tired of this, he brought it to the desk and was given P–M 2, the time being noted as before (see Appendix E). In many cases the entire 70 to 80 minutes devoted to this test was consumed before the second test was returned, so that only the first score was exactly comparable for all cases. Both were retained, however, for the value they might be for such cases as turned in P–M 2 before time was up.

The number test, which was fortunately given first, was not

P-M 1

PROBLEMS

Name..... School.....

Date..... Grade.....

Magic Number Squares

Here is a Magic Square:

8	1	6	=15
3	5	7	=15
4	9	2	=15
15	15	15	15

DIRECTIONS

Notice that the numbers in the magic square at the left add up to 15 in all possible directions—up and down, across and across corners.

Use as many of the practice squares below as you need to in working out another magic square. Put a different number in each box so that it will add up to the same amount in all directions—up and down, across and across corners. When you have made a magic square in any of the practice squares then copy it carefully in the square marked FINAL SOLUTION.

PRACTICE—

1. First Trial
2. Second Trial
3. Third Trial
4. Fourth Trial
5. Fifth Trial
6. Sixth Trial
7. Seventh Trial
8. Eighth Trial
9. Ninth Trial
10. Tenth Trial

FINAL SOLUTION

FIGURE 67

MAGIC NUMBER SQUARE OR P–M 1 TEST

P-M 2

PROBLEMS

Name..... School.....

Date..... Grade.....

Magic Word Squares

Here is a Magic Square

E	A	S	T
A	V	E	R
S	E	M	I
T	R	I	P

DIRECTIONS:

In the square at the left you will see that the letters are so arranged as to spell just four words, and that each of the four words can be read from top to bottom, as well as from left to right.

You are to try to find other combinations of four-letter words which will make a similar Magic Word Square. Remember the same word must appear twice; once across, and once from top to bottom, and there may not be more than four words.

You may use as many of the Practice Squares below for trials, as you need. When you have found a solution, put it in the square "FINAL SOLUTION". Then if you wish to, go ahead and try to find a second solution, and put the answer in the square marked "2nd final solution".

PRACTICE—

1. First Trial
2. Second Trial
3. Third Trial
4. Fourth Trial
5. Fifth Trial
6. Sixth Trial
7. Seventh Trial
8. Eighth Trial
9. Ninth Trial
10. Tenth Trial
1st FINAL SOLUTION
2nd FINAL SOLUTION

FIGURE 68

MAGIC WORD SQUARE OR P-M 2 TEST

solved by any of our subjects in 75 minutes. But the word test used as an illustration a series which suggested an easy solution to a good many pupils, and other solutions were discovered, thus vitiating the results for some 200 cases. As already noted, however, the second test was not used in our total score, so that our main results were not affected.

The Persistence for Self and Persistence for Class Tests. These two tests are derived from the results of the Efficiency Coöperation test described in Chapter III. There were two sets of scores for each child, representing speed of work on monotonous simple additions. One set of scores was for six units of work done for self and the other set was for six units done for class. As these twelve units of speed work are continuous, without any rest period, there was a tendency for fatigue or ennui to set in and reduce the speed of work. This was particularly true of the last four minutes of the twenty-four, during which the amount of work done per minute dropped considerably. There were, of course, wide individual differences in the ability to resist the tendency to go more and more slowly. This ability we have classified as "persistence."

As a measure of persistence in this sense, we took the differences between the work done in the first two periods (four minutes) and the work done in the last two periods, subtracting the first score from the second. A large negative score thus means a big falling off or lack of persistence. As the differences approach zero and become positive they represent increasing persistence. In dealing with the units of work done for self, we called the score "persistence for self," and in dealing with the units of work done for class we called it "persistence for class."

The amount done in the first four minutes compared with the amount done in the last four minutes gives a rough indication of the pupil's fatigability. It may go up like this:

or down, like this:

or remain on one level, like this:

14977

The actual data for two typical cases are given in Table XL.

TABLE XL

TYPICAL ILLUSTRATIONS OF AVERAGE NUMBER OF EXAMPLES DONE FOR SELF AND FOR CLASS IN THE FIRST TWO AND LAST TWO PERIODS OF THE EFFICIENCY COÖPERATION TEST

	CASE A		CASE B	
	For Self	For Class	For Self	For Class
First two periods	50	60	70	65
Last two periods	75	55	50	40
Derived persistence scores	+ 25	− 5	− 20	− 25

From this table it may be seen that case A was more persistent than case B both for self and class, and that each was more persistent for self than for class. Case A, however, showed a difference of 30 points (+ 25 minus − 5) between his two persistence scores, being far more selfish *relatively* than B, who showed a corresponding difference of 5. In absolute terms, however, A worked harder for class than B did, *viz.*, − 5 against − 25.

In order to combine the two persistence scores into one, we simply added them, their SD's being equal. But their significance lies largely in the differential effect of the two motives, self and class, on the tendency to persist against fatigue and loss of interest.

The exact directions for the persistence tests have not been quoted in the text since the inhibition tests, about to be described, were administered at the same time in the case of population C, and in later groups other combinations were used which altered the procedure. In Appendix E, however, we have printed the directions as they were used in later testing. These should be read carefully in order to make clear the type of experience involved in the tests.

This coucludes the description of our five tests of persistence, and we shall now turn to the other aspect of self-control, which we have called inhibition.

INHIBITION

As we have just stated, the classroom inhibition tests were first administered along with the tests of persistence. Six techniques were tried out in population C:

1. The Story Inhibition test
2. The Safe Manipulation test
3. The Puzzle Manipulation test
4. The Ruggles Distraction test
5. The Picture Inhibition test
6. The Candy Inhibition test

These roughly classify in two manipulation tests (2 and 3), two distraction tests (4 and 5), and two competing desires tests (1 and 6), thus roughly measuring three types of inhibition with two tests each. As there were two stories in the first test, seven articles in the third, and twelve distracting pages in the fifth, besides twenty examples in the Ruggles test, the total number of temptations was large enough to give a fair sampling of this sort of behavior.

The Story Inhibition Test. Two procedures were used. In one each child holds a copy of the story, which is read aloud by the examiner to the climax. The examiner then says:

> I want you to guess how the story will come out. Turn the sheets over and write on the back of the last page how you think the story will come out. It would be a fine thing for you to think it out by yourself, but if you feel that you have to see what really happens you may break open the last page (which had been previously pasted to the preceding page) and read the rest of the story. (The last sentence is repeated.)

Here the drive is to know how the story ends and the child is asked to see if he can inhibit it and, instead, guess how the story comes out. If he chooses to guess, he is not told how it comes out.

The other procedure is a little different in that it sets one drive off against another. The two drives used are the desire to know the end of the story and the desire to make a score in a test. All pages of the story are stuck together, showing only the first page. Each child is given a copy and the examiner reads the first page aloud to the children. At the bottom of the page are these directions:

If you want to make a good score, go back and count all the letters in each line and write the number on the short line at the right before turning the page. The first line is already counted. If you don't care about your score, you may of course break open the next page and read the story. Take your choice. There is not time for both test and story.

The child is told that the score will not count on his monthly report. The choice, however, is clear cut. He may either finish the story or go back and count letters, which would yield a letter-counting score.

The score on this test is whether the subject broke open the one sealed page of the first story or one or both of the sealed pages of the second story, giving a range of 0–3.

The original forms were mimeographed. We had experimented with a great variety of techniques aimed at measuring the pull of the story interest against the desire for a score, such as rate of counting letters or of underlining A's while reading interesting material as compared with rate of counting letters or underlining A's in uninteresting material, but had found nothing that would give a well-distributed score without very large errors of measurement. In one class, for example, the children not only improved their rate of routine work on the interesting material but read the story at the same time with sufficient understanding to retell it afterwards. We were obliged to resort to some such mechanical device, therefore, as we have just described, which gave an unambiguous record of the failure to resist the tendency to read the story in the interest of a good mark on the test — or, in the other case, the preference of finding out how the story came out to getting a score on the test.

Our mimeographed forms worked very well, but when we had the stories printed for later use, in spite of our efforts to get opaque paper, the material actually delivered was so transparent that a pupil could read through the closed page by holding the test to the light, and in the second story the third page could be read by holding the second closely to it. As it was too late to make changes, we gave the tests and checked on whether the pupils actually did hold them to the light. The results of the second story only were used, and only the first break was counted. Those who broke the seal were scored 0, and those who did not were scored 3. The following quotation will illustrate the technique, which is described fully in Appendix E.

Copyright 1927, Institute of Educational Research, Teachers College, Columbia University.

S-I-2

Name ..... Date .....

Grade ..... School .....

DIRECTIONS: If you want to make a good score on this test, **do one page at a time.** The pages are stuck together. Leave them alone until ready to turn the page.

The Mystery of the Big Box
OR
An Adventure With A Spook

Begin here:

Bobby was playing about the railroad station, thinking	[46]
about a mysterious robbery that had taken place. Lots of	[ ]
stamps and money had disappeared without a single door	[ ]
or window being opened. "If I could only catch that rob-	[ ]
ber!" said Bobby to himself.	[ ]
After a time Bobby wandered into the freight room.	[ ]
It was rather dark and spooky in here. Rats and mice	[ ]
scampered about. There was an odd, musty smell. There	[ ]
was one game Bobby liked to play in here. The darkness	[ ]
made it more exciting. This game was a hurdle race with	[ ]
himself. He would place the various sized packages here	[ ]
and there and jump over them in turn.	[ ]
Today he found he could have a queer sort of hurdle—	[ ]
a big box that was very heavy and had an undertaker's label	[ ]
on it. Bobby had a feeling that he might be doing wrong	[ ]
to use the coffin to jump over but he went ahead and used	[ ]
the long box as his best hurdle.	[ ]
Suddenly, in almost his first attempt to jump over the	[ ]
coffin, he stumbled. With some force he fell against it.	[ ]
Then he felt his teeth chattering! For something had dis-	[ ]
tinctly said, "Oh!"	[ ]

(Continued on the next page)

Now:—If you want to make a good score, go back and count all the letters in each line and write the number in the brackets at the right before turning the page. The first line is already counted.

If you don't care about your score you may of course break open the next page and read the story. Take your choice. There is not time for both test and story.

The Safe Manipulation Test. At the beginning of a long test period there was placed on each pupil's desk a small toy safe with a combination lock set at $6\frac{1}{2}$. The following directions were given:

> I am going to place on each desk a toy safe, like this (holding one up) and a puzzle like this* (holding one up). These are to be used in a test. I am going to ask you not to touch either of these now because, a little later, you will be given a chance to see if you can put the puzzle together and open the safe. If you touch them now, that will give you an unfair advantage over the others. This is why I would rather you would not touch them until time to begin the test.

After the safes were passed out the examiner proceeded to give a series of paper-and-pencil tests. After each test the papers were collected and at the same time the examiner checked the dial of each safe, noting on each child's paper any changes that might have been made up to that time. The more papers there were passed out and collected, the more opportunities there were to check on the safes. Six such collections and check-ups were made in the final procedure. When ample opportunity had been given to everyone to manipulate the safes and after six checks had been made, the directions were given to take the safes and see how many could work the lock and get it open. A sheet of paper was provided on which the successive motions necessary for opening the lock were to be recorded, and to stimulate interest in the procedure we placed a penny in each safe, which was to become the property of those who succeeded in getting inside. (See Appendix E.)

The score on this test should be the length of time the child manipulated the safe when he was asked not to touch it. We could not secure this, but the number of times the dial was in a different position on the six inspections gives a good measure of the tendency to yield to the desire aroused. In using these records we called six changes 0, five 1, four 2, etc., in order to have higher scores represent greater inhibition.

The Puzzle Manipulation Test. In our measurement of deception we had occasion to use as one of our puzzle tests a game called

---

* This was the wooden cross puzzle used to measure persistence.

Puzzle Peg, which consisted of a board with small holes in it of the accompanying pattern. Every hole but one contains a peg. This board was in a box. In this box we placed five small puzzles, carefully arranging them among the pegs, some of which were removed for the purpose, so that afterward they could be put back where they were only by following a master pattern.

These lures were passed out with the following directions:

I am going to pass out all our things together. The box (holding it up) has several things in it which we would rather you would not disturb. If you get these things disarranged, it means a lot of work for someone to put them right again. We shall use the boxes presently.

These boxes were left on the desk while the pupils took a series of paper-and-pencil tests. After about thirty minutes a test form was passed having on it a design like that in the box. The examiner then gave the following directions:*

This sheet has a design on it just like the one in your box. Look at the box but don't touch the objects in it. Notice that most of the holes have pegs in them. Be sure the box has the side where it says T away from you. Now take your paper and put an X on every circle on your paper to show just where the pegs are, and leave alone the other circles. Don't move the things to see where the holes are. Just put an X where the pegs are. Go ahead.

After three minutes the papers were collected and the pupils were once more warned not to touch the puzzles. The boxes were then collected and taken to another room and inspected, scored, and made ready for another group. A numbering device enabled us to identify each box.

Scoring the Puzzle Manipulation Test. We found upon investigation that the puzzles placed in the Peg Board box had fairly definite degrees of attractiveness, so that those who touched the less attractive also touched the more attractive. The scale was

* In schools X, Y, and Z a score sheet was prepared for this test and the deception test, which was used at the same time. There were no essential changes in the directions. See Appendix F.

not perfect by any means, but the steps were reasonably clear. In order of appeal the objects handled were as shown in Table XLI, beginning with the least attractive.

TABLE XLI

RELATIVE ATTRACTIVENESS OF OBJECTS IN THE PUZZLE MANIPULATION TEST

OBJECTS	PER CENT TOUCHING	PER CENT NOT TOUCHING
1. Pegs	9	91
2. The whole board	19	81
3. A bent nail puzzle	19	81
4. A spring and ring puzzle	22	78
5. A trick nut and bolt puzzle	28	72
6. A coin and leather puzzle	34	66
7. A moving balls puzzle	39	61
Any object		59

From the figures of Table XLI we determined a rough rating for the objects, assuming that the absolute zero of inhibition was about as far below the power to resist moving a peg, which 9% failed to show, as this was below moving the whole board, which 19% failed to show. Each object not touched was given a score of 1, with 9 as the score of those who did not touch anything and 0 as the score of those who touched everything.*

In working out our results with this technique at school C, which will be reported in Chapter XX and in Appendix F, we did not use this "amount" score, but only a "fact" score, giving a child a mark of 0 if one or more objects were touched and 1 if nothing was touched.

* The per cents of Table XLI were translated into their sigma values as for the deception scale described in Chapter XII, Book Two, of *Studies in Deceit*. Thus the σ values of the per cents failing to inhibit are -1.3, $-.9$, $-.9$, $-.8$, $-.5$, $-.4$, $-.3$, $+.5$, assuming that we take as the score for those who touched nothing the sigma value of one-half 41%. If $-3.0\,\sigma$ is placed at zero, -1.3 becomes 17 tenth-sigmas and the other scores, respectively, 21, 21, 22, 25, 26, 27, 35. These slightly modified, and the zero shifted to a point

Copyright, 1927, Institute of Educational Research
Teachers College, Columbia University, New York
Devised by Allen M. Ruggles

FORM C-2

Begin here:
7 2 3 4 5 3 7 8 6 1 3 5=
9 1 7 6 8 4 3 1 2 8 4 3=
6 3 2 1 5 4 2 5 7 2 3 7=
3 7 4 3 7 5 3 6 8 3 1 5=
9 1 5 4 3 2 7 5 4 8 9 4=
7 3 6 1 5 7 8 3 2 8 4 3 6=
2 6 3 5 3 2 4 5 7 1 2 4=
3 6 4 1 3 6 2 3 4 2 7 1=
8 2 7 5 2 3 4 5 9 6 4 2=
5 2 3 4 1 3 5 4 7 2 1 7=
5 4 6 4 2 5 6 1 3 4 5 3=
2 5 3 2 4 8 6 5 1 2 5 4=
6 4 5 3 2 1 4 6 7 8 3 2=
7 2 4 5 6 1 3 7 6 4 3 2=
4 5 1 2 9 4 3 6 5 6 3 2=
3 4 2 4 8 7 6 5 1 4 5 4=
5 4 6 1 2 5 3 4 1 3 5 4=
3 4 5 8 9 4 3 2 1 6 3 2=
4 6 5 3 4 2 5 8 6 1 2 5=
9 1 7 6 5 4 3 1 4 6 5 7=
Peter Piper picked a
NO
Oh DEAR!
PICKFORD
Candy
ICE CREAM
The earth is flat. It can be proved.
Hang on
Siqaret
Hong Gong
BEGIN OVER
Why don't you—
CANDY
THIS SPACE RE SERVED
I AM SICK
FIRE
Circus
ITS PINK
CAT
SPARROWS
THIS ALSO
BANG
WHO ?
Hot Dogs
WARNING
Bonjour
Jerusalem
Cows
Horses
Sheep
Children
Goats
Skunks
SKANA-
BREAD
LARD
Gwonkiskwischcki
Iowa, U.S.A.
A YANKEE SETTLEMENT
NUT & JEFF

FIGURE 69

FACSIMILE OF THE RUGGLES DISTRACTION TEST

The Ruggles Distraction Test. The sheet represented by Figure 69 was devised by Allen M. Ruggles to measure distraction. His technique required that the digits of each line be added and subtracted alternately, as shown by the directions quoted here:

Copyright, 1927, Institute of Educational Research, Teachers College, Columbia University, New York
Devised by Allen M. Ruggles

FORM C-1b

Name..... Date.....

School..... Grade.....

ARITHMETIC TEST

On the lower half of this sheet are twelve rows of figures. Each row is a separate problem which you are to work as follows:

Example **8 2 4 3 4 1 5 6 3 2 4 7=9**

To the first figure add the second, from this sum subtract the third, then add the fourth, subtract the fifth, add the next, subtract the next, etc., placing at the end of each line of figures the final result which you get.

In the problem above, the first figure is 8. Then 8 plus 2 is 10; 10 minus 4 is 6, plus 3 is 9, minus 4 is 5, plus 1 is 6, minus 5 is 1, plus 6 is 7, minus 3 is 4, plus 2 is 6, minus 4 is 2, plus 7 is 9, which is the answer.

---

as far below 17 as this is below the next step, we have 0, 3, 6, 6, 6, 9, 12, 12, 21. Divided by 3, these become 0, 1, 2, 2, 2, 3, 4, 4, 7.

This gives the following scoring key:

TABLE XLII

SCORING KEY FOR THE PUZZLE MANIPULATION TEST

	LEVEL OF INHIBITION	SCORE STEP
No inhibition	0	0
Peg	1	1
Board, or nails, or spring	2	1 1 1
Bolt	3	1
Coin, or balls	4	1 1
Not represented	5	
Not represented	6	
Nothing touched	7	3
Maximum score		7 + 2 + 1 = 10

Each object not touched thus counted 1 on an inhibition score, with 10 for the score of those who did not touch anything. In actual practice, we used 9 as the value of the top level, thus giving a bonus of 2 points instead of 3 to those who touched nothing.

Do the same thing with the problems below, taking the ones down the left hand side of the page first before beginning the ones on the right hand side.

WHILE WORKING THE PROBLEMS DO NOT PLACE YOUR HANDS OR ARMS ON THE PAPER. You are not to use your pencil in any way in following the figures. Be perfectly quiet and make no sound with your lips as you are working. When you have finished a problem write the answer quickly at the end of the line after the sign = and then take your hands off the paper and place them on the desk at the sides of the sheet or under your desk.

Work as fast as you can and still be accurate.

PROBLEMS

9 1 3 2 5 4 6 7 6 7 5 1= 5 4 3 1 6 7 4 3 2 5 4 2=

1 5 4 8 7 3 2 6 8 5 4 7= 8 1 2 3 4 2 7 8 1 2 9 5=

Twelve examples were given to familiarize the pupils with the procedure, and then twenty examples without any distracting pictures were given in order to measure the pupils' ability with this type of arithmetical computation. The examples in all cases used the same digits as those of the test itself, but with sufficient variation in order to make either memorizing or copying impossible. Five minutes were allowed by us for practice and six minutes each for the undistracting and distracting sheets.

This procedure we at first simplified by having the pupils only add the digits rather than add and subtract alternately, and our scores for school C are based on this method, which is described in the following directions:

Copyright, 1927, Institute of Educational Research, Teachers College, Columbia University, New York
Devised by Allen M. Ruggles

FORM C-1a

ARITHMETIC TEST

Name.....Date.....

School.....Grade.....

On the lower half of the sheet are twelve rows of figures. Each row is a separate problem in additions.

Example: 1 3 6 7 2 4 1 9 5 8 6 2=54

To the first figure add the second, to this sum add the third, etc., placing at the end of each line of figures the final result which you get. In the problem above 1 plus 3 is 4, plus 6 is 10, plus 7 is 17, plus 2 is 19, plus 4 is 23, plus 1 is 24, plus 9 is 33, plus 5 is 38, plus 8 is 46, plus 6 is 52, plus 2 is 54, which is the answer.

Do the same with the problems below, taking the ones down the left hand side of the page first before beginning the ones on the right side.

WHILE WORKING THE PROBLEMS DO NOT PLACE YOUR HANDS OR ARMS ON THE PAPER. You are not to use your pencil in any way in following the figures. Be perfectly quiet and make no sound with your lips as you are working. When you have finished a problem write the answer quickly at the end of the line after the sign = and then take your hands off the paper and place them on the desk at the sides of the sheet or under your desk.

Work as fast as you can and still be accurate.

PROBLEMS

9 1 3 2 5 4 6 7 6 7 5 1=	5 4 3 1 6 7 4 3 2 5 4 2=
1 5 4 8 7 3 2 6 8 5 4 7=	8 1 2 3 4 2 7 8 1 2 9 5=

We found, however, that this made the task too easy for pupils in the seventh and eighth grades; so we used it in grades 5 and 6 and retained the more difficult procedure in grades 7 and 8.

Through faulty administration, the results of the testing in populations X and Z had to be thrown out, and later the Y results were also found to be useless as measures of inhibition since the pupils did about as well with the examples when the figures were surrounded by the distracting pictures as the standard group did when no pictures were used. This test, therefore, does not appear in our later results. We have nevertheless printed this description for the benefit of others who may wish to experiment further with this technique.

Scoring the Ruggles Test. The Ruggles test calls for a double testing technique, similar to that used with the IER and Speed tests of deception, and therefore our procedure in standardizing and scoring it was exactly comparable to the procedure described fully in *Studies in Deceit.** Briefly, the score containing a measure of inhibition or its opposite, distractibility, was the difference between the pupil's performance for the first six minutes and his performance for the six minutes when the distracting features were present. Any falling off in amount of accurate work done could be attributed to the failure to inhibit the tendency to look at the scrawls and screeds surrounding the examples. As our first rough indication of distractibility we therefore used the difference between the number of examples

* Pages 65 to 82, Book One.

done in the first six minutes and the number of examples done when the pictures were present.

A more accurate measure would allow for normal variation from one six-minute period to another. This expected variation was found by testing another population of similar age-grouping and ability, only using in this case, in addition to the forepractice, two six-minute periods without any distracting features for the second period. We then found the average gain from the first period to the second and the standard deviation of all the differences between the two periods. As a pupil might normally be expected to gain, any loss should be regarded as a loss not from the first score, but from the first score plus the average gain. We used the SD of the undistracted differences as our unit of measure, dividing each difference exhibited by our populations by this figure to give an Xi inhibition score.

The actual figures used in determining these basic facts are given in Tables XLIII–A and XLIII–B.

TABLE XLIII–A

STANDARDIZATION DATA FOR THE RUGGLES TEST

GRADES	N	MEAN GAIN	SD OF DIFFERENCES
5, 6	139	.34	2.35
7, 8	115	2.80	2.60

The figures of Table XLIII–A gave us two formulas for computing the scores on the Ruggles test. Calling the score on the first form of the test C1a for grades 5 and 6 and C1b for grades 7 and 8, and on the second form C 2, the same form being used for both groups, the score for 5 and 6 then becomes

$$\text{Inhibition Xi} = \frac{\text{Ca} - \text{C1a} - .34}{2.35},$$

and for grades 7 and 8

$$\text{Inhibition Xi} = \frac{\text{C2} - \text{C1b} - 2.8}{2.6}.$$

After the Ruggles tests were scored in this way we found that the mean inhibition Xi for population Y was $-.51$, which is only $\frac{1}{2}$ Xi below 0, or the mean of the standard group. Also the SD of Y was only 1.43 Xi as against an SD of 1.00 for the standard group, showing that the pictures

did not greatly increase variability. Only 11 out of 273 children, furthermore, had Xi scores in excess of 3.00.

Considering the upper and lower grades separately, the facts are as follows:

TABLE XLIII–B

	N	Mean Number of Examples Gained	SD of Gain	$\frac{SD}{\sqrt{N}}$	Difference ÷ σ Difference
Standard 5 and 6 . .	139	.34	2.35	.20	5.5
Population Y 5 and 6 .	159	− 1.32	3.08	.24	
Standard 7 and 8 . .	115	2.80	2.60	.24	1.33
Population Y 7 and 8 .	114	2.20	4.03	.38	

The significant difference in the case of grades 5 and 6 was found to be due to two groups of boys.

Finally, the correlation between this test and three out of the four other inhibition tests is negative, is − .12 with the Check List, and .06 with total reputation for inhibition.

The Picture Inhibition Test. This is similar to the Ruggles test, except that we used simple additions for the activity, knowing from our work on the Speed tests of deception that this material gave highly reliable results in one or two minutes of work. As in the use of additions for measuring deception, we first determined the subject's normal speed of work. For this we gave four minutes of practice in two-minute intervals. Then we passed out a booklet containing not only the same addition problems but also, at the top of the pages, a good deal of distracting material. (Full directions will be found in Appendix E.) The directions for the second booklet began as follows:

> The next folders must not be opened until you are told to open them. When I say, "Go," write your name, school, grade, and the date on the lines provided. Ready? Go! (Allow time for this to be done.)
>
> Stop. Pencils up. Now before you turn the page let me explain that on the following pages are some pictures, stories, and other interesting things, as well as some more additions problems. You will have time to look at them for only a short time. If you want to make a good

score, follow directions. Start adding promptly when I say, "Go," and stop when I say, "Stop." Turn to page 4. Here is a story about a boy who got lost in a dark forest. (Examiner reads first three lines aloud.) You may have a few seconds to read while resting your hand. Keep pencils up. (Allow 20 seconds.)

All ready now. If you want to make a good score, do the additions of Test 1 as fast as you can. Begin at the left. Go. (Allow 2 minutes.)

Stop. Pencils up. Turn at once to page 6. This is a garage. There are twelve mistakes in the picture, and on the right is a picture puzzle. There is a third boy hidden in the picture. Keep your pencils up while looking at the picture and rest your hand. (Allow 45 seconds.)

All ready now. If you want to make a good score, do the additions of Test 2 as fast as you can. Begin at the left as before. Go.

There were six exercises in all, and the remaining distractions consisted of a pair of sensational news headlines about some boys caught in a burning house and a girl who might prove to be a real princess; some riddles and their answers in pictorial form, such as," Who is the shortest man in the Bible?" and as the answer on the opposite page, a picture of a shoe and a man standing alongside; several nuts to crack and pictorial answers, such as, "How much dirt is there in a hole 1 foot long, 1 foot wide, and 1 foot deep?" and the answer, a picture of an empty hole in the ground; and for the sixth exercise a comic strip picturing a schoolroom and the conventional visitors asking fool questions. Figure 70 will illustrate the appearance of one of the twelve pages used in this part of the test.

The assumption underlying this test is the same as that of the Ruggles test. If a child fails to get to work on time, or allows his eyes to wander back to the picture before time is up, his score will fall below the amount done on the practice trial, where there were no pictures. We found it necessary to get the pupils actually looking at the distracting features, since many pupils in our preliminary trials with this technique would simply cover up the pictures with one hand so as to avoid the temptation of looking at them at all. Since we must have something to inhibit if we are to measure inhibition, we did our best to start going an interest in the

pictures and also to keep alive the more general interest in getting a good score. The matter was stated in such a way as to make the

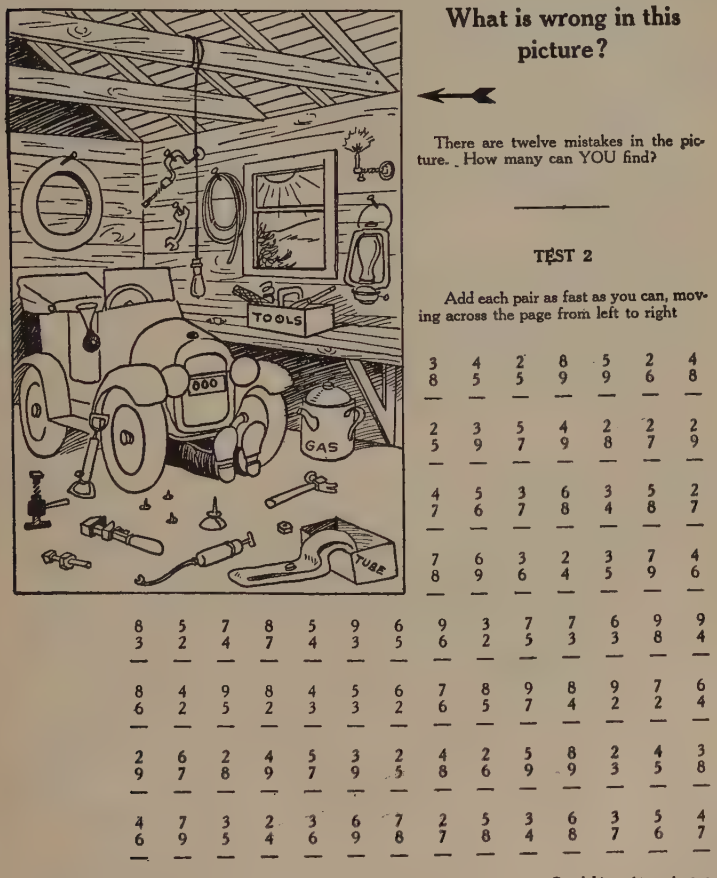


FIGURE 70

ILLUSTRATION OF THE PICTURE INHIBITION TEST *

*Copyright, 1927, Institute of Educational Research, Teachers College, Columbia University, New York City.

taking of the test the approved activity and the distraction the disapproved activity.

In standardizing the material of the Picture Inhibition test, we had twelve pages of additions set up to correspond exactly to the pages of the test except that no pictures were printed. The preliminary practice folder was also used, just as for the regular test. Again, the differences between the scores without distraction and the scores with distraction contained the measure of inhibition. The undistracted normal gains and the normal variability of a standard population are given in Appendix C. Briefly, a loss of 78 or more examples on the six trials when the distracting features were present indicates statistically that the chances are 999 or more in 1000 that the loss was due to the failure to inhibit the tendency to look at the pictures, etc. while doing the examples. The standard population, working without the pictures, showed no losses greater than 45, but if we were to give an actual failure score to an individual, we should prefer to overstate the case in order to be fair to him.

The Candy Inhibition Test. As an advertisement, a large candy concern* furnished us with several hundred small sample boxes of candy at a nominal cost of five cents a box. Each box contained six pieces — two chocolate peppermints, two salted nuts, and two hard candies. At the beginning of the last testing period at school C these boxes were passed out with the candy exposed to full view, and the following statement (see Appendix E) was made:

> As a reward for taking so many tests we are going to treat everyone to a small box of candy. This candy is yours, but we ask that you leave it in the boxes until we have finished the exercises for to-day. If everyone is chewing candy, it adds to the confusion. But we will pass it out now anyway.

The candy remained on the desk for forty minutes while a series of paper-and-pencil tests was taken. At the end of each test, as the examiner collected the paper, he placed a figure on the paper to represent the number of pieces that had been removed. As the

* Page and Shaw.

order in which the papers were used was constant, one could thus tell not only how much was taken, but just when each piece was taken. Hence it was possible to compare the pupils with regard to the promptness with which they yielded to the temptation to eat the candy before the time was up as well as with regard to the fact of their yielding and the total amount eaten. Our crude scores consisted of the bare fact of having taken out at least one piece of candy during the forty minutes.

Difficulties in the administration of this test led us to drop it, which was perhaps a mistake, since in its results it turned out to be one of the best tried.

THE TOTAL SELF-CONTROL SCORES

As in the case of our service tests, we made various combinations of our persistence and inhibition tests in spite of their specificity, and in presenting our results we shall use both these totals and the scores on the separate tests.

Persistence Totals. Our first subtotal, which we may call A, consisted of the sum of the scores on the Cross and Ring test and the Magic Square test. Each score was stated in terms of minutes of work, and so the two were added as they stood. The second subtotal, called B, consisted of the sum of the Persistence for Self and the Persistence for Class tests, the sigmas of which were equal. As a third subtotal, we took the Story Resistance score alone.

We then made distributions of these three subtotals and found their respective sigmas, being thus able to give them equivalent weightings. We then added the weighted scores. As there were two tests in the A and B subtotals and but one in C, we doubled the value of A and B as compared with C.*

Inhibition Totals. The Picture test was treated as one unit and the sum of the Story Inhibition, Puzzle Manipulation, and Safe Manipulation

* The respective sigmas were for A, 24; B, 52; and C, 13, or in the ratio 2, 4, 1. Their weightings inversely would be $\frac{1}{2}$, $\frac{1}{4}$, 1; combining these with the suggested logical weighting of 2, 2, 1, we have a final weighting of 1, $\frac{1}{2}$, 1, or 2, 1, 2. The original raw Story Resistance score was the sum of two decimal ratios, which for convenience was multiplied by ten; that is, the original SD was 1.3 instead of 13 for this test.

as the other. In order to combine the last three, the score on the Safes was reversed (as already noted) so that larger scores would signify greater inhibition. These scores ranged from 0 to 6. Only the second Story Inhibition test was used, the first having been spoiled by poor administration in the case of several rooms. We gave a score of 0 to those who broke the seals to read the remainder of the story and a score of 3 to those who did not. Giving the Puzzles a weight of 3, the Safes a weight of 2, and the Stories a weight of 1, we were thus enabled to add the three raw scores as they stood for the subtotal inhibition score.

The grand total inhibition score was found by giving about equal weights to the Picture test and to the subtotal just described and adding them, first so changing the Picture scores as to make them all positive. The original scores run from about minus 50 to plus 50. We grouped the scores in fives, calling the lowest group 0 and adding 1 to each successive group of five scores, giving a range from 0 to 20.*

ADMINISTRATION OF SELF-CONTROL TESTS

In school C, where these techniques were given their first try-out as a single battery, we set up a program of testing lasting several days. The way in which this program was carried out constituted the test situation and must therefore be described in some detail.

The sequences were as follows:

First Day. The Story Resistance test, for measuring the tendency to persist in overcoming a mechanical difficulty in reading to find out the end of a story which had been read to the dramatic point.

Second Day. Safe Manipulation and Ruggles Distraction tests and Puzzle Mastery tests, for measuring inhibition of the tendency to handle an interesting object and the tendency to look at funny scrawls and screeds instead of working, and, in the case of the Japanese wooden cross, for measuring the tendency to persist at a difficult task.

The safes were passed out first, with the explanation already quoted. Then the Ruggles test was given, as described, with exactly two minutes allowed for passing out the first sheet. After the test the papers were collected by two examiners, the position of the safe dial being indicated on

* The SD of the subtotal is 3.7 and of the Picture test 15.3. By telescoping the Picture scale from 100 to 20, we approximately equated the two sets of scores and so added them directly. In dealing with the Picture test by itself, we have used the raw differences.

the paper as it was collected. Four minutes were allowed for this and for passing out the next sheet, which was the actual Ruggles test. Again the safe dials were checked as the papers were collected.

Then a series of letter counting exercises was given, with two minutes allowed for each of four sets of similar exercises, and four minutes between them.

Thus there were six successive periods of four minutes each when the pupils had nothing to do but look at the safes. That is, they had to inhibit the desire to touch the safe for twenty-four minutes. The temptation was somewhat aggravated by the routine character of the last exercises.

At the end the pupils were allowed a brief time to solve the combination lock and open the safes.

At the close of these exercises the wooden crosses were explained and passed and the time each pupil spent on this puzzle was noted.

Third Day. The Puzzle Manipulation test and paper-and-pencil mastery problems, for measuring the inhibition of the tendency to handle attractive objects and persistence at a difficult task.

The puzzle box was passed out with its accompanying test paper and the Travelers' Problem already described. These were numbered to match concealed numbers on the boxes so that the boxes could later be identified. The boxes, containing five attractive puzzles, were supposed to be left untouched while the mental problems were attacked. As the Travelers' Problem was solved or abandoned, the Chain, River, and Magic Square puzzles were given out in this order, the time devoted to each being noted. Then at the end of the period of 30 minutes, the pupils were asked to mark the sheets having on them the Peg Board design to show where each peg was (not an easy thing to do without yielding to the temptation to move the objects). When these papers had been passed in, the boxes were collected.

Fourth Day. The Candy Inhibition test, the Story Inhibition test, and the Picture Inhibition test, for measuring the power to inhibit the tendencies to eat the candy, to finish the story, and to look at the pictures, puzzles, and jokes.

The candy was passed out first. Then the two stories which could be finished only by breaking open the pages were used, and finally the two folders of the Picture Inhibition test. Thus four papers were collected, the amount of candy taken being checked up on each.

The Fillers. In order to keep pupils occupied with uniform tasks when they completed any of the puzzles or problems on which they were working

ad libitum, we provided additional material of a similar nature to use as fillers. These were particularly needed in connection with the persistence tests. Some of these consisted of routine letter counting sheets, of which we had quantities made for this purpose, and some consisted of other mental puzzles. These were worked in as circumstances served to suggest. We shall discuss later the possible effect on pupils of seeing those who stop trying go up and get something else to work at.

Subsequent Use of Inhibition and Persistence Tests. In our later testing, to be described more fully in Volume III of the report, we combined our techniques in such a way as to leave the persistence aspect of the test free from complication with the inhibition aspect of another test.* This required additional innocuous fillers and a relatively longer period for each test. The new order was as follows:

Testing Session	*Test*
2d	Picture Inhibition
3d	Persistence for Self and for Class
4th	Story Persistence or S–M, with puzzle fillers
6th	Magic Squares or P–M, with puzzle fillers
8th	Cross and Ring, with puzzle fillers
11th	Puzzle Manipulation, with a mechanical deception test and letter counting fillers
12th	Safe Manipulation, with a mechanical deception test and letter counting fillers
16th	Story Inhibition test

The other sessions were used for a large variety of tests to be discussed in the third volume.

* See Appendix E for directions concerning the self-control tests.

CHAPTER XIX

REPUTATION FOR SELF-CONTROL

In accordance with the theory discussed in Book One, we felt it wise to find out all we could about the behavior of our subjects quite apart from our test program. Although our main interest in gathering this material was to provide a total reputation score, many of the items used referred to the forms of behavior we were incorporating in our tests. These we lifted from their context and used as an indication of reputation for the several conduct tendencies measured. In the case of inhibition we have opinions of both teachers and pupils, but no persistence items were included in the " Guess Who " technique. All the devices are described in Chapter IV, so that all we need report at this point are the particular items used and their statistical value.

REPUTATION FOR PERSISTENCE

On the Check List there are nine words which it was felt might have some bearing on this tendency. These are as follows:

Quitter	Disloyal	Unreliable
Steady	Persistent	Impulsive
Faithful	Shirker	Constant

Each child was supposed to be marked on both forms of the Check List by two persons. As the words were divided between Form A and Form B, we took each sheet as a unit, giving the subject a +, −, or 0 for the sheet according as the rater checked words indicating persistence or the reverse, or simply did not check him at all, thus placing him automatically in the center of the scale. These scores then ranged from + 4 to − 4.

Only three items of the Conduct Record referred to persistence. These are quoted here:

1. Persistence
   (*a*) Works as long as others or if the work is not laborious (0)
   (*b*) Sticks to a task until finished even if disagreeable or involving real sacrifice (+ 2)
   (*c*) Gives up as soon as any difficulty arises (− 2)
   (*d*) Quits before others do (− 1)
   (*e*) Works longer than others (+ 1)
2. Industriousness
   (*a*) Very active, but does not hurt himself working (+ 1)
   (*b*) Shirker and slacker — lazy (− 2)
   (*c*) Hard worker; loves to be busy (+ 2)
   (*d*) Works along without much urging (0)
   (*e*) Works when urged; loves his ease (− 1)
3. Control of interest and attention
   (*a*) Works consistently fast in spite of fatigue (+ 2)
   (*b*) Tires quickly of a job; wants to do something else (− 2)
   (*c*) Warms up in the beginning, but soon weakens (− 1)
   (*d*) Fairly consistent in his work and interest (0)
   (*e*) Works consistently for some time until much fatigued (+ 1)

The Conduct Record items were roughly scaled from + 2 for the item representing greatest persistence to − 2 for the item representing least persistence. The third item was marked independently of the others and was given more attention by the raters. It was therefore given double weight. This yielded a range for the three items of + 8 to − 8. The scale values are given above with the paragraphs describing each step on each scale. The midpoint was found by noting the frequency with which the raters checked each paragraph. Ordinarily there was a clear middle point representing the paragraph most frequently used. This was called zero. On each side of this neutral position there were two of the remaining four paragraphs under each item, which were scored + 1 or + 2 and − 1 or − 2 according to the frequency of use. Thus a child who was checked in the square for (*b*) of item 3 was one of only a very few who fell in this group, and he was consequently scored − 2. But a child who was checked (*d*) on the

same item was scored neutral or zero, as he was one of many who were not distinguishable as either persistent or its opposite.

In combining the two sets of records, the scores were added as they stood, a procedure which, on account of the wider range of the Conduct Record scores, gave the latter double weight. This we felt was justified by the greater reliability* of this device. The estimated reliability of this total reputation score is .85.

Validity of the Reputation for Persistence. As we have no opinions from the pupils concerning one another in the matter of persistence, we can give as evidence of validity only the correlation of one teacher's ratings with another's. This is .576 for Check List with Conduct Record and .48 for the Check List alone. The theoretical validity of .86 shows that there is a considerable area of pupils' conduct which is not common to the experience of different teachers. The pupils, further, whose opinions we do not have, would see one another in situations to which the teachers do not have access. We have thus an incomplete measure of reputation for persistence.

REPUTATION FOR INHIBITION

For this we have pupils' opinions as well as teachers'. The latter are recorded on the Check Lists and Conduct Record sheets as usual. The following words were selected from the Check List as representing the power of inhibition or the lack of it:

Morally weak	Wasteful	Spendthrift
Dissipated	Deliberate	Lawless
Rowdy	Hot-headed	Temperate
Imprudent	Impulsive	Cool
Foolhardy	Unthinking	Reserved
Provident	Impetuous	Prudent

* Two items, by a fourfold table, yielded a correlation of .579. For two against two, this rises to .71, and assuming the third to be of the same order of reliability, for three against three, .80.

The Check List reliability was computed by a sixfold table, placing the zeros in the center. This gave a product moment r of .41 for Form A against Form B, which becomes .58 for A and B combined and .64 for all four sheets combined, by the formula given in the footnote on page 93.

Injudicious	Conscientious	Disorderly
Cautious	Reckless	Scatterbrained
Dutiful	Thoughtful	Forbearing
Thrifty	Unreflecting	Economical

As before, there were four sheets for each child, and each sheet was marked +, −, or 0 according as the pupil was checked with positive words, with negative words, or not at all. All four sheets thus added to a total ranging from + 4 to − 4.

In the case of the Conduct Record we formulated eight items to stand expressly for this phase of conduct. These were as follows:

1. Control of temper
   (*a*) Impulsive, but does not have violent outbursts (− 1)
   (*b*) Will stand for a lot, but gets angry after a while and shows it (0)
   (*c*) Cool, calm, and balanced at all times (+ 1)
   (*d*) Hot-headed, flies off very easily (− 2)
   (*e*) Slow to anger and hard to irritate (+ 2)
2. Control of interest and attention
   (*a*) Works consistently fast in spite of fatigue (+ 2)
   (*b*) Tires quickly of a job; wants to do something else (− 2)
   (*c*) Warms up in the beginning, but soon weakens (− 1)
   (*d*) Fairly consistent in his work and interest (0)
   (*e*) Works consistently for some time until much fatigued (+ 1)
3. Control of laughter
   (*a*) Able to keep a perfectly straight face always (+ 3)
   (*b*) No restraint; bursts right out any time (− 1)
   (*c*) Controls except when others are laughing (+ 1)
   (*d*) Restrains outbursts, but smiles or grins (+ 2)
   (*e*) Shows restraint only under pressure (0)
4. Control of tendency to talk too much
   (*a*) A regular chatterbox, but is quiet in school (− 2)
   (*b*) No restraint; talks constantly unless checked (− 3)
   (*c*) Talks a good deal, but knows when to stop (− 1)
   (*d*) Very guarded in speech (+ 1)
   (*e*) Answers questions and talks freely, but only when asked (0)
5. Control of tendency to tell a secret
   (*a*) Will tell a few friends unless he has given his word of honor
   (*b*) Just simply cannot keep anything to himself
   (*c*) Will not tell unless asked; then he cannot refrain

(*d*) Can keep only secrets of great importance when telling would injure someone
(*e*) Close-lipped; never tells once he has given his promise

6. Control of desires
   (*a*) Wants what he wants at once and howls or makes a fuss if he doesn't get it (− 2)
   (*b*) Can indefinitely postpone any desire (+ 2)
   (*c*) Can wait a while but not long (0)
   (*d*) Doesn't seem to want things he can't have (+ 1)
   (*e*) Wants everything he sees others have and cannot rest until he gets them (− 1)
7. Control of appetite
   (*a*) Never refuses sweets, but knows when he has enough
   (*b*) Allows himself only so much and never exceeds it
   (*c*) Once he gets started, he cannot stop
   (*d*) Overindulges in sweets; eats all he can get
   (*e*) Can have candy before him and never touch it
8. Control of body movements
   (*a*) Fidgety, noisy, and continually moving (− 2)
   (*b*) Can sit perfectly still for long periods (+ 2)
   (*c*) Remains quiet for a while, but soon gets restless and squirms around (− 1)
   (*d*) Makes little noise; usually quiet (+ 1)
   (*e*) Quiet only when interested in work (0)

Numbers 5 and 7 were omitted because of lack of data, and 2 was retained although also used in the reputation score for persistence. Following the practice just now described, the zero point of each scale was determined from the results obtained when the device was used. The scale values are printed along with the steps of the scale. In two instances, the neutral point was apparently not in the middle, and so the scores ranged from + 3 to − 1 or − 3 to + 1.

In finding the reputation for inhibition which the pupils had among their own classmates we used the following items from the "Guess Who" device, the general nature of which is described on pages 87 to 91.

5. This is someone who controls his temper and never gets angry.
9. Here is someone who always wants his own way. This person wants to be boss and have everyone do as he says.

14. Here is one who is not annoyed when things go wrong. He is happy and cheerful and calm, no matter what happens.

17. Here is someone who eats more than is good for him, especially sweet things, and either can't or won't stop.

20. This is someone who is decent and clean in all his conversation.

21. Here is someone who flares up and gets mad on the slightest excuse.

23. This is someone who controls his appetite and can always resist candy and sodas and gum when he knows he ought not to eat them.

24. Here is one who tells disgusting stories and uses profane language.

Numbers 5, 14, 20, and 23 are positive, indicating the ability to resist temptation; and 9, 17, 21, and 24 are negative, indicating weakness. Each child has the opportunity of naming as many of his classmates as he chooses for the distinction of being one who controls his temper, or, *per contra*, who flares up and gets mad on the slightest excuse, or who possibly eats more than is good for him.

As we said before, these items were not prepared with the purpose in view of using them as measures of reputation for possessing the power of inhibition, and in selecting these particular paragraphs we have probably erred on the side of too broad an interpretation of the meaning of the term. Certainly we have included a wide range of behavior not represented in the tests, so that any failure on the part of our tests to correspond to such reputation as may be indicated by the pupils' opinions need not be wondered at.

The scores on the "Guess Who" test ranged from + 12 to − 12, on the Check List from + 4 to − 4, and on the Conduct Record sheets from + 12 to − 12. In combining these three sets of scores, the pupils' opinions were regarded as equal in significance to the sum of the two sets of teachers' opinions, and the Conduct Record was given $1\frac{1}{2}$ times as much weight as the Check List.*

Reliability of the Reputation for Inhibition. On the assumption that the reliability of the "Guess Who" items dealing with inhibition would be about the same as we found it to be for the service

* The respective sigmas are: Conduct Record, 3.8; Check List, 2.8; "Guess Who," 5.0. As the logical weights were 3, 2, and 5, this made it possible to get the weighted sum by adding the raw scores as they stood.

items,* the predicted coefficient of reliability for the eight paragraphs dealing with inhibition is .86. In the case of the Check Lists, the correlation of the scores given by the same teacher on the two forms, which were filled out at different times, is .561, which yields a reliability coefficient of .76 for all four sheets (two teachers).† In the case of the Conduct Record device, items 1, 2, and 3 were paired off against items 4, 6, and 8. The resulting coefficient for the entire six is .82.

The combined reputation for inhibition, representing the opinions of both teachers and pupils, is quite satisfactory for the ground covered by these devices. If we should go back another time and get similar records from the same teachers and pupils about the same children, it is to be expected that the two sets of measures would correlate about .90; that is, our two sets of scores would agree closely.

Validity of the Reputation for Inhibition. The reader may use his own judgment in deciding whether the items included in the reputation score for inhibition really belong together in a picture of this type of behavior. The teachers do not agree with one another nor with the pupils in expressing opinions by the means we have described.† It is probable, therefore, that each rater is thinking of situations in which he has seen the subject but in which the others have not seen him. That is, in order to cover an individual's entire range of temptations, one would need considerably more data than we have collected. If we had nine more such measures as we now have, we should feel that we would know the subject's reputation about as well as if we had observations of him in a very large number of situations.‡

---

* See page 93.

† The r between the Check Lists of two teachers for the same pupils is .52. See formula on page 93.

‡ The inter-r's of the several reputation devices are Check List and Conduct Record, .631; Check List and "Guess Who," .258; Conduct Record and "Guess Who," .475; averaging .455, which for three such devices gives a predicted theoretical validity (Formula (13), *Studies in Deceit*) of .715. Upon crossing the pupils' opinion with the teachers' by combining the Conduct Record and Check List in one score, the validity becomes .763.

CHAPTER XX

VALUE OF THE CLASSROOM TESTS OF SELF-CONTROL

NOTE: The general reader may prefer to turn at once to the concluding section, on page 341.

As with our tests of service tendencies, we shall report any facts that may throw light on the significance of our measures of self-control. Our first interest is in the reliability of the tests. Our second interest is in the extent to which the test results correspond to general reputation for self-control. Our third interest is in the extent to which a theoretical total set of measures of self-control is represented in the tests we have used.

THE PERSISTENCE TESTS

RELIABILITY

Our evidence for the reliability of these tests is drawn solely from populations X, Y, and Z.* These three populations are widely different in social structure, X being a fairly exclusive, well-to-do community, Z a foreign and rather poor community, and Y a small town. The reliabilities are reported for all three groups combined.

The Story Resistance Tests. These tests, it will be recalled, begin by interesting the pupils in a story and then making it difficult for them to find out how the story ends. To read beyond the point of suspense, it is necessary to plow through material that is run together, with capitals and small letters misplaced and

* For descriptions of these populations, see the Introduction and also Chapter IX.

words improperly spaced. The score on each story is the ratio of the number of words actually read in the story to the number read in three minutes on standard material prepared in the same way. Since there were three stories, we had ordinarily two scores. The last story could not be counted since time was up before most of the children had done as much as they would have done if time had been unlimited. Indeed, there were some pupils whose scores for the second story were unusable for the same reason. We had enough cases (782), however, to give us a reliable coefficient of correlation between the score on the first story and the score on the second. This r is .595, which gives a predicted coefficient of reliability of .75 for the two combined. Considering the variation in story material, this is quite as high as one would expect.

The Cross and Ring and Magic Square Tests. In these tests, as the reader will remember, the task was to solve a difficult puzzle, taking as much time as was desired. The pupils were allowed to quit and take another puzzle whenever they felt like it. The scores are amounts of time spent on the first puzzles attempted.

We could not take the correlation between the time spent on the first mechanical puzzle — whether Cross or Ring — and the second — whether Ring or Cross — as the total time was too short to allow a large proportion of pupils to get tired of the second puzzle. Consequently any r based on all cases would tend to be zero or negative, since the longer they spent on the first puzzle the shorter the time left for the second. If we took only those cases that quit on both, they would be a selected group possessing relatively less of the tendency being measured than the rest, which would render the correlation ambiguous. The two paper-and-pencil puzzles were so unlike as to make their correlation of no value as an indication of reliability, and here again those who spent most time on the Magic Number Squares had less time for the Magic Word Squares. So we cannot measure the reliability of either test.*

The Persistence for Self and Persistence for Class Tests. These scores were derived from the Efficiency Coöperation test,

* The time spent on the first mechanical puzzle correlates .38 with the time spent on the first paper-and-pencil puzzle.

which required constant adding of simple digit combinations for twenty-four minutes, twelve under the motive of gain for self and twelve under the motive of gain for one's class. The decline in speed as fatigue set in measured the reverse of the tendency to persist. As the work was done in twelve units, one half could readily be correlated with the other half. When working for self, this r is .79. When working for class, it is .85. These yield reliability coefficients of .88 and .92 respectively.*

Total Persistence Score. By combining the tests into two batteries of four tests each, a rough measure of the reliability of the total persistence score was readily obtained. The two batteries consisted of A, the first Story Resistance test, the Cross and Ring, half of Persistence for Self, and half of Persistence for Class; and B, the second Story Resistance test, the Magic Number Square, half of Persistence for Self, and half of Persistence for Class. The r between these two combinations is .799, which yields a predicted reliability coefficient for the total persistence score of .89. That is, if we should repeat the present set of tests, the r between the two sets of scores would be .89, or closely similar. This places the persistence tests on the same level of reliability as those dealing with the other forms of behavior we have studied.

Correspondence between Conduct and Reputation

We have described in the preceding chapter the methods adopted for measuring the reputation of our subjects in the matter of per-

* These various r's may well be summarized at this point.

TABLE XLIV

Reliabilities of Persistence Tests

Test	Observed r	Prophesied Reliability
Story Resistance	.59	.75
Cross and Ring with Magic Number Square	.38	.55
Persistence for Self	.79	.88
Persistence for Class	.85	.92
Whole battery	.80	.89

sistence. We must now point out the extent to which this reputation agrees with the results of our tests.

First, it should be stated once more that the reputation scores are based only on the opinions of teachers, so that there is a narrower basis for judgment than in the case of inhibition and service. This is somewhat offset by the fact that the behavior incorporated in the tests occurs only in the classroom. If the situations and responses included in the reputation scores were identical with those included in the tests, this narrowing of opinion might be of advantage to a clear understanding of the correspondence to be expected between tests and reputation; but examination of the items actually available as indicators of reputation for persistence will show that the opinions are far more general and cover a far wider area of experience than the tests.

With this introductory explanation, we shall not be surprised to find that there is only a slight agreement between the total persistence score and the total reputation for persistence. The Story Resistance and Magic Number Square tests are somewhat better indicators of reputation than the others, but the correlation for the total is only .23.*

These statements, of course, concern the entire population of X, Y, and Z combined. If we leave out of consideration the middle group of cases, who are not as well measured as the rest, we find a

* The complete table of r's is as follows:

TABLE XLV

CORRELATIONS BETWEEN TESTS AND REPUTATIONS FOR PERSISTENCE

TEST	CHECK LIST	CONDUCT RECORD	TOTAL REPUTATION
Story Resistance	.33	.12	.25
Magic Number Square	.26	.12	.21
Cross and Ring	.00	.05	.08
Persistence for Self	.10	.10	.12
Persistence for Class	.13	.12	.14
Total	.20	.17	.23

much closer correspondence between what the pupils do and what they have the reputation for doing. For instance, four out of five of those among the 25% most persistent, as measured by the tests, are regarded by the teachers as more persistent than not.

Theoretical Validity

Here we deal with the capacity of the tests already constructed to stand for a very large number of similar tests — or with the whole range of persistent behavior among children. We have used only five, and these have a validity of .78.* We would need twenty-four more to have enough to represent all the persistent tendencies of our cases. But it should not be forgotten that our tests as they stand are reasonably accurate measures of the particular conduct tendencies which they do measure.

True Validity

As already indicated, we do not have any pupils' opinions regarding the persistence tendencies of their classmates; but if we should assume as much agreement between teachers and pupils as in the case of other behaviors, *viz.*, about .40, we can figure what we have called the true validity of these tests. The general discussion of the meaning of this method of validation is given in Book One. It signifies roughly the correlation between our tests, when free from

* The intercorrelations on which the validity coefficient is based are given in Table XLVI.

TABLE XLVI

Intercorrelation of Persistence Tests — Population XYZ

Test	2	3	4	5
1. Story Resistance	.35	.16	.15	.18
2. Magic Number Square		.38	.20	.15
3. Cross and Ring			.22	.18
4. Persistence for Self				.42
5. Persistence for Class				

Average inter-*r*, .239

errors of measurement, and the opinions of teachers and pupils, when these are free from errors of measurement and prejudice. The coefficient is only .52.* In view of the difference between the subject matter of the tests and of the opinions, which can readily be seen by referring to Chapter XIX, this low validity is not surprising.

THE INHIBITION TESTS

Reliability

Considering first the question of reliability, the problem is to determine how well these tests measure anything. We shall consider later how well they measure inhibition. The tests were planned to go in pairs so as to secure two measures from each of the four types of situations studied: (1) the presentation of a story to its moment of greatest suspense and the introduction of a task requiring the pupil to leave the story at that point (the test was to see whether or not the pupil could inhibit the desire to know the ending); (2) the presentation of small attractive objects, to test how well the pupils could inhibit the tendency to handle them; (3) the presentation of addition examples together with visual distractions, to test the effects of the distraction on speed of adding; and (4) the presentation of a small box of nuts and candy to see if the pupil could refrain from eating them for an hour. Each situation except (4) is covered by two tests and (4) was used only in population C. If each pair of tests can be regarded as alike in the sense of requiring the same type of performance, then the correla-

---

$$* \text{TrV} = \frac{r_{tc}}{\sqrt{R_t}\sqrt{R_c}} = \frac{.30}{\sqrt{.59}\sqrt{.57}} = .52.$$

The denominator is found, as before, from the average intercorrelation of the five tests, .239, and the r between teachers' and pupils' opinions, which is taken as .40.

We estimate that the r between the persistence total and a combination of pupils' and teachers' ratings would be about .30, since the r between persistence and the teachers' opinions alone is .23. This is on the assumption that the r would rise in the same proportion as in the case of reputation for service and for inhibition. Hence .30 instead of .23 is used in the numerator.

tion between the scores of each pair will give some indication of how reliably each situation is tested.

The classroom inhibition tests were tried out in population C before being administered to our major populations X, Y, and Z. We have, therefore, a check on the statistical values of these instruments. In discussing each test, we shall quote from both experiments but shall give the detailed consideration of the results of the work in population C in Appendix F.

The Story Inhibition Test. In population XYZ we have 612 cases who took both story tests. The correlation between the first and second story, each scored 0 or 1, is .48.* In school C we get .46. In view of the fact that this test admits only of a dichotomous score, these reliabilities seem satisfactory. If one story correlates with another story .48, then two stories with two more would correlate .65. As we reported in Chapter XVIII, the results for one story could not be used in the case of several groups. Hence we have but one left for the whole population and two for about 600 children.

The Manipulation Tests. In school C we found that, when certain groups used for experimenting with the administration of these tests are omitted, the r between the Safes test and Puzzle Manipulation test is .50. In population XYZ the r is only .318 when all groups are included, but again, by excluding certain groups in which there seems to be some doubt as to whether the examiner followed directions carefully, the correlation goes up to .496. We feel reasonably sure that, if these two tests are properly given, they will correlate about .50 with each other. If two more equally good tests were devised and given, the r between two and two would be about .67.

The Visual Distraction Tests. In school C we found a correlation of .42 between the Ruggles test and our Picture Inhibition test. As already stated, the Ruggles test, unfortunately, was spoiled in giving in all but population Y, and even here it failed to work satisfactorily. This leaves the Picture test as the only one of this type. But it is in six parts or periods of two minutes each

* Pearson's ϕ.

and offers a good opportunity for split-form correlation. In school C we found that the odd-numbered exercises correlated .90 with the even-numbered. The data on population X, Y, and Z yield a corresponding correlation of .80. This gives it a total reliability coefficient of .89.

To summarize, then, the reliability of the Story Inhibition test is .65, of the Manipulation tests, .67, and of the Visual Distraction tests, .89.

In order to determine the reliability of the entire battery, two sub-batteries were formed. For population C these consisted of the following sets of tests:

Sub-battery A	Sub-battery B
The odds of the Picture test	The evens of the Picture test
The first Story test	The second Story test
The Safes test	The Puzzle Inhibition test
The Ruggles test	The Candy test

The correlation of sub-battery A with sub-battery B is .628, which gives a coefficient of reliability of .77 in population C for the entire series.

For population XYZ the same combinations were made except that the Ruggles and Candy tests were omitted. The resulting coefficient is .80. This is lower than in the case of our other behavior tests, but such r's are likely to differ on account of the fact that the sub-batteries can never be made equivalent.

Correspondence between Conduct and Reputation

As in the case of the persistence tests, our measures of behavior do not cover the same ground as our measures of reputation. We merely pulled out of our devices for finding reputation in general such items as seemed most likely to represent what we call inhibition. The Picture Inhibition test agrees most closely, but the correspondence of all is only slightly better than in the case of persistence.* The agreement of the whole battery of tests with the total reputation scores is expressed by the r of .40.

* All the r's are given in Table XLVII (at bottom of next page).

In addition to the correlations for the entire population, which of course include all the neutral or colorless cases, we shall consider the extremes to see the extent to which the reputation scores agree here with the tests. First we should see the general nature of the distribution of test scores and reputation scores. These are given in Figures 71 and 72. Unlike the curve for service scores, which are concentrated in the middle, these curves are more normal. Nevertheless, let us take from one end of the reputation scale all rated − 10 or lower and from the other end all rated + 12 or higher. This gives about 25% of all the cases. Now how are these found on the various tests? We shall illustrate by reference to the Picture Inhibition test.

Let us call all whose Picture score was − 1.05 Xi to + 3.00 Xi gainers, or +, and all whose score was − 1.05 Xi to − 4.5 Xi losers, or − . Also we shall call all rated − 10 or lower, −, and all rated + 12 or higher, +. If we put these four groups in a diagram, it would look like this:

	− in reputation	+ in reputation	Total
+ on test	23	56	79
− on test	53	41	94
Total	76	97	173

% + + − − 63.5

If the 79 of the upper part of the table and the 94 of the lower part

TABLE XLVII

CORRELATIONS† BETWEEN INHIBITION TESTS AND REPUTATION

TEST	*r*
Picture Inhibition . .	.385
Puzzle Manipulation .	.234
Safe Manipulation . .	.232
Story Inhibition . . .	.122
Total	.400

† These are biserial *r*'s except the first, which is a product-moment *r* corrected for errors of type I as discussed in *Studies in Deceit*, Book Two, Chapter VI.

and the 76 of the left side and the 97 of the right side had been arranged just by chance, the numbers in the cells would have been as follows:

	−	+	
+	35	44	79
−	41	53	94
	76	97	173

% + + − − 49.0

The difference between the per cent falling by chance in the perfectly correlated sections, the ++−−, and the per cent actually found is 63.5 minus 49, or 14.5.

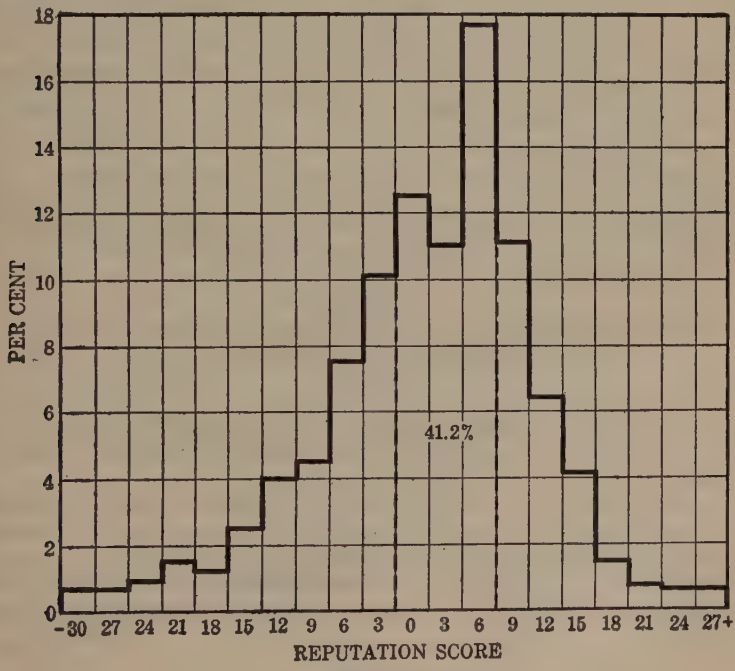


FIGURE 71

PERCENTAGE DISTRIBUTION OF THE INHIBITION REPUTATION SCORES — ALL POPULATIONS COMBINED

N = 808

In a similar way the differences between the per cents actually falling in the ++ −− cells and those expected by chance are: for the Safes (those touching *versus* those not touching), 12%; for the Puzzles, 7.3%; and for the Stories, 3.0%. Thus the Picture test agrees best with reputation. The other differences, though less extreme, are significant.*

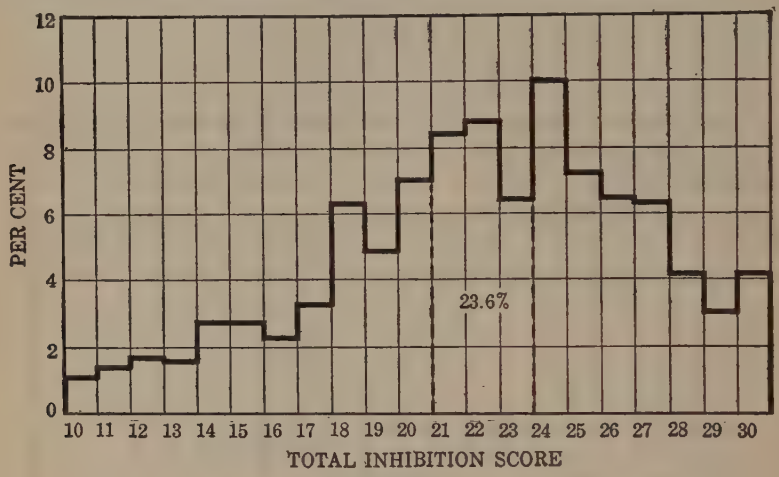


FIGURE 172

PERCENTAGE DISTRIBUTION OF THE TOTAL INHIBITION SCORES — ALL POPULATIONS COMBINED

N = 640

* Since 49 is the per cent of −− and ++ that would be found among these proportions by chance, any deviation from it indicates the presence of differentiating factors.

However, the 49 is not absolutely stable but varies with different samples. The cause of such variation is the difference in proportions. For example, when the four total N's around the margin of the table are all the same, then the chance of ++−− is always .50. But when these N's are not the same, the chance per cent of ++−− changes. Suppose an extreme situation like this:

	+	−	
−		1	1
+	99		99
	99	1	100

When the whole population is included in these fourfold tables, dividing the pupils roughly into those who inhibit and those who do not on each test, and those who are reputed to be well controlled and those who are reputed not to be, we find that the actual proportion of those marked + on both counts and those marked − on both counts is far greater than would occur by chance.*

Theoretical Validity

By this is meant the capacity of the set of tests already developed to stand for a much larger hypothetical set, covering the entire range of behavior of the sort represented in the tests.† This capacity is represented by an r of .65. It would take 47 tests, or 43 more than we have, to cover the ground completely.

Before leaving the problem of validation, it should be reported that in population C the theoretical validity of the six tests used

---

Now the chance per cent with such proportions is 98.03, while the actual per cent is 100. Hence the probable error of the difference between the chance per cent of ++−− cases and the actual per cent of ++−− cases depends on the probability that any given proportion of N's would recur. Slight differences in the proportion will not materially affect the probability that the ++−− group will be 50% of the total.

* Except for the Story Inhibition test. The r's for the several fourfold tables are: Picture Inhibition, .423; Puzzle Manipulation, .466; Safes, .458; Story Inhibition, .150.

† And assuming, of course, an average intercorrelation equal to that actually found. The inter-r's are given in Table XLVIII.

TABLE XLVIII

Intercorrelation of Inhibition Tests — Population XYZ

	2	3	4
1. Pictures	.12	.16	.13
2. Puzzles		.32	.12
3. Safes			.11
4. Stories			

Average, .16

was .805, and omitting the Ruggles test, .818.* Some eighteen more would have covered the ground for population C. We are inclined to believe that this is a truer picture of the actual validity of the inhibition tests than is represented in the results of population XYZ.

True Validity

We have two approaches to the pupils' reputation for inhibition, one from the standpoint of the teachers and the other from the standpoint of the pupils. It may be, as in the case of the service tests, that, when we consider what is common to tests, teachers' impressions, and pupils' impressions, we shall find here a closer agreement between behavior and reputation. And in fact this is the case, for the teachers' opinions correlate .41 with the pupils', indicating a fairly small amount of common experience, and we can determine the agreement at the point where opinions and tests strike common ground. This so-called "true validity" proves to be .80,† which, while of only theoretical value, does suggest the probability that if reputation and test dealt with the same areas of experience they would agree closely.

* The inter-r's for population C are given in the following table:

TABLE XLIX

Intercorrelation of Inhibition Tests — Population C

	2	3	4	5	6
1. Stories	.277	.073	.320	.050	.210
2. Safes		.500	.256	.001	.384
3. Puzzles			.307	.177	.325
4. Pictures				.420	.240
5. Ruggles					.010
6. Candy					

Average, .237

Average, omitting Ruggles, .289

† $\mathrm{TrV} = \frac{r_{tc}}{\sqrt{R_c}\sqrt{R_t}} = \frac{.40}{\sqrt{.433}\sqrt{.582}} = .80.$ See pages 112 to 117.

CONCLUSIONS CONCERNING THE VALUE OF THE CEI MEASURES OF SELF-CONTROL

It appears that our tests of persistence and inhibition represent about one-fifth of the situations which should be embodied in any single battery purporting to measure an individual completely in these modes of conduct. The battery of five persistence tests has a reliability of .89, and each of its parts gives reasonably consistent results. The battery of from four to six inhibition tests has a reliability of .80, and except for the Picture test the parts are not so reliable as those of the persistence battery.

The agreement between reputation and tests is as great as could be expected, in view of the heterogeneous nature of the situations covered by tests and opinions, and it would seem probable that a high degree of correspondence between reputation and conduct would be found if devices were used which related definitely to the same facts.

CHAPTER XXI

INDIVIDUAL TESTS OF SELF-CONTROL*

There are many situations in which it seems desirable to test inhibition, persistence, and interests that are not adapted to group administration. We have worked out a series of such tests, which are given individually and in a private room.

PERSISTENCE

The Pennies and Ball Bearings Test — Hunting Objects. The aim of this test was to see how long the subject would search the floor for pennies and ball bearings presumably dropped by the examiner. A certain room had been set aside for the use of the examiner. The balls and pennies were placed about the room according to a predetermined plan, before the entrance of each subject. The formula used was: "Now, X, I've just dropped some things on the floor. I'll tell you what they are — some steel balls and some pennies. Will you please pick them up and put them in this box?" Box lay on center table. Examiner set stop watch and went on about work, apparently ignoring the child. If subject asked, "Is that all?" examiner said: "I don't know. Is it?" Time until subject definitely gave up was recorded. Score is the time spent and the total number of objects found.

Standing on Right Foot. The aim was to see how long the subject would stand on one foot. The subject was told: "Now I want to see how long you can stand on your right foot, without losing your balance and without touching your other foot to the floor. Understand? Is your right foot all right?" If subject showed

* These tests were developed and administered with the assistance of Miss Ruth E. Welty, from whose report this chapter is drawn.

tendency to rest his left foot against his leg, he was cautioned to let leg hang free. The time was recorded.

The Cracker Whistling Test. The examiner said to the subject, "Can you whistle?" He was asked to show how he could whistle. "I want you to show me how quickly you can eat this cracker and whistle. Here is a cup of water which will help you to whistle sooner, but I would rather you did not take any of it. Eat the cracker over the waste basket, so that the crumbs fall into it." Examiner made effort to tell when the cracker was eaten and to record time, but this proved difficult as mastication could not be distinguished from saliva-stimulating movements. The score is the time from the moment the cracker is taken until the subject succeeds in whistling.

The Dime Puzzle Test. The examiner gave the child a puzzle consisting of two pieces of metal, with a dime visible which could not be obtained until the puzzle had been solved. Dime puzzle held before subject. "Do you see this? It has a dime in it. If you can get the dime out, you may have it." Examiner set about doing other work. As the first puzzle became so worn with handling that solution became easy, another type of puzzle was substituted after the twelfth subject. As the purpose of the experiment was to tell how long the child would remain at a baffling task when there was a possibility of definite reward, the change of type of puzzle halfway through the experiment was deemed allowable. The score is the time elapsed before the subject gave up.

INHIBITION

A Candy Test. This is similar to the group test already described. Six pieces of candy in an open box were placed on the table and the examiner said, "Now this candy is yours; I would rather you didn't eat it now, but you may if you wish." The examiner then proceeded with the program. The score is the number of pieces eaten before the end of the series of tests.

Mirror and Feather Test — Tickle Test. Examiner asked subject: "Do you know what a wooden Indian face is? A face that

doesn't express anything, just as if it were made of wood. Understand? (Explanation continued until subject says he understands.) I want to see if you can make your face a perfect wooden Indian face. Remember not to show anything, whatever happens to you. Look at your face in this mirror, and I will look too." This gave examiner a chance to go back of child. "Is your face fixed just the way you want it to be?" After answer "Yes" or "No," examiner tickled the back of subject's neck with a feather. There were three slow passages from ear to ear. Reaction recorded in record book. No reaction is scored 0; any perceptible facial reaction is scored 1; maximum score is 3. The aim of this test is to see if the child can inhibit any tendency to react overtly to this stimulus.

Picture Book Test — Laughter Test. A scrapbook containing 53 pages of funny pictures was placed before the child. The pictures were selected to cover a wide range of types of humorous pictures. The following directions were given: "You like to look at pictures, don't you? Here are some pictures. I want you to look at every one of them, and still keep your wooden Indian face." If subject started to go rapidly through the pages, he was cautioned, "Go slowly and look at every one of them." The hand mirror was placed on the table in such a way that examiner could look into it and see subject's expression continuously. The score is (1) the time taken in looking and (2) the number of laughs or smiles — each laugh counting 2 points and each smile 1 point.

The Bad Odor Test. A malodorous drug (assafoetida or in some cases sulphureted potash) was used as the stimulus. The examiner handed the subject an uncorked bottle containing the assafoetida or the open box containing the potash and said: "Hold this close under your nose and take three long slow smells of it. Remember to keep a wooden Indian face." Any perceptible change of facial expression was scored 1 point; maximum score 3 points.

The Bad Taste Test. A drug with an unpleasant taste, such as castor oil, unrefined cod liver oil, argyrol, was placed on the tongue of the subject. The examiner proceeded as follows: Subject shown bottle and asked, "Do you know what this is?" Castor

oil used only in the first six cases. After that, shifts were made between the two other drugs. In these later cases, if the subject said, " Castor oil," he was assured that it was not. Examiner said: " I'm going to put two drops of this on your tongue. Then you close your mouth and keep a perfectly good wooden Indian face for two minutes. It can't hurt you. If you have to spit it out, use this towel." Subject closely observed for two minutes. The score is the degree of reaction. Three degrees were recorded, which were scored 1, 2, or 3.

The Startler Test. The examiner used a small instrument called a " startler," which makes a grating, rasping noise and emits sparks but is entirely harmless. This device is placed about six inches in front of the child's face, and he is told, " Look straight at this, and keep a wooden Indian face." The lever is depressed; the noise is made and the sparks fly. The score is the strength or violence of the responses, scored in levels 0, 1, 2, or 3.

The Pain Test. The aim of this test is to determine the subject's tendency to avoid pain, or power to resist the tendency. Subject is seated before the Whipple pain balance apparatus, a device consisting of a lever which presses down on a nail of the finger as weights are added at the tip of the long fulcrum of the lever. What is actually taking place is more or less successfully concealed from the subject by a screen. Subject is told: " This will not hurt you in the least. Place your right hand here, and your first finger so. Now I want you to tell me when your finger first begins to feel uncomfortable, and I'll stop." Weights added until subject gives desired reaction. Then lever lifted and subject told: " Now put it back. I'm going to add more weights, and I want you to tell me when you don't want to stand any more of it."

The number of grams used before the subject reported that pain was felt was first recorded. Then the number of grams added to this before the subject indicated a desire to stop was recorded. The score is roughly the difference between the pain weight (the point where pain is first felt) and the quitting weight.

SCORING THE INDIVIDUAL TESTS

Three of the tasks yielded a time score, which was the number of minutes until the task was abandoned. They were:

1. Hunting balls and pennies
2. Standing on right foot
3. Working at dime puzzle

The whistling-after-cracker task also yielded a time score, but it is not comparable with the foregoing, since it involves effort to shorten the time required, plus the ability to forego the help available from drinking water. In the present treatment of the data this task will be disregarded.

Six of the tasks yielded results in the form of kind and degree of reaction. They were:

1. Feather	tactual stimulus
2. Humor book	visual-emotional stimulus
3. Malodorous drug	olfactory stimulus
4. Drug with bad taste	taste stimulus
5. Startler	combined visual-auditory stimulus
6. Candy	visual-taste stimulus

Study of the records of the subjects in these six tasks revealed the fact that the reaction could be stated in finer terms than a yes-no record. That is, there were four comparatively well-marked categories into which the reactions could be placed. Since the perfect performance of the task required inhibition of any response, the appearance of any sort of response was considered more important than the kind of response. Degree of response, then, is the factor upon which is based the weights assigned to the several responses.

Weight	*Reaction*
0	No response; complete inhibition of observable reactions
1	Faint response; partial inhibition of observable reactions, as a smile in the humor situation, a frown or brow-puckering in the disagreeable situations, blinking in the startling situation
2	Full response; complete performance of the normal reaction to be expected in the situation when no instruction forbidding expression has been given

3 Extreme response; exaggerated performance of the normal reaction to be expected from the situation, as an explosion of laughter in the humorous situation, nausea in the disagreeable situations, laughter plus sharp withdrawal in the startling situation

Situation (6), refraining from eating the candy, has 0 assigned to complete abstinence and 3 assigned to the eating of all the pieces.

It must be stated that the performance called "normal reaction," in weights 2 and 3 above, is a common-sense application and therefore subjective. It is based, however, on the accepted observation that under normal conditions the average person when tickled will smile or wriggle, when shown a humorous picture will smile, when confronted with substances of bad odor and taste will express aversion, when startled will change his position or expression in some way. Whether or not the person who under normal conditions fails to show these reactions really feels them and inhibits them or actually fails to make the connection between the stimulus and the response is no part of the present study.

VALUE OF THE INDIVIDUAL TESTS

The subjects in the experiment were 25 in number, 12 girls and 13 boys. They were all residents of an orphanage of the better class, housing a total of fifty children. The age range was from seven years and ten months to fourteen years and eight months, and the IQ range was from 78 to 131. The median IQ was 92, the average 90.4. All the children had been given the Stanford-Binet tests within a period of a month preceding the experiments, by the experimenter herself.

The results of the experiments — the behavior of the child in each of several situations — were recorded by the examiner at the moment of happening. The diary method was used wherever possible. The exact reaction was set down and later evaluated.

This method of recording data, since it is done by one person, the observer, is open to the criticism of inexactness and unverifiability. In reference to the first objection it may be stated that only a rough measure of the reaction was wanted. Response in

terms of observable human behavior was the thing being studied. Minute variations in degree of reaction, as well as invisible and inaudible reactions, were in the present instance disregarded. Emphasis was placed on observable, socially influential behavior.

In regard to the second point, unverifiability, it is admitted that to the degree that there is no check-up on the results the method is lacking in one of the essentials of scientific method. However, if there be taken into account the approach to the experiment, which was that of careful, exact inquiry into the reactions to certain applied stimuli, recorded immediately in terms of observable muscular and verbal reactions, and the fact that the results are offered as such — simply the reactions of the subject at that particular moment — the point of unverifiability will have been given due weight.

The significance of these individual tests, given here in a preliminary and crude form, may be judged by the distribution of the scores, their intercorrelation, the correlation of each with a total of the remainder, and the correlation of the total with an external measure. These facts will be briefly reported for both the persistence and inhibition tests.

The Tests of Persistence. The three tests of persistence yielded the distributions in Table L.

These distributions are reasonably satisfactory for so small a number of cases. The intercorrelations of the tests are also of considerable interest. The times spent on the three tasks do not at all agree. There is practically no relation between persistence in standing on one foot, persistence in hunting for the balls and pennies, and persistence in solving the puzzle in order to get the dime.* Consequently, the relation between the time spent on

* These inter-r's (rank difference) are as follows:

	Balls and Pennies	Working on Puzzle
Standing on one foot	− .108	+ .100
Balls and Pennies		+ .088

TABLE L

DISTRIBUTIONS OF TIME SCORES OF INDIVIDUAL PERSISTENCE TESTS

HUNTING BALLS AND PENNIES		STANDING ON RIGHT FOOT		WORKING ON DIME PUZZLE		TOTAL PERSISTENCE	
Minutes	Cases	Minutes	Cases	Minutes	Cases	Minutes	Cases
Over 10	1	15	1	40	2	60	2
10		14	1	35	1	55	1
9		13		30		50	
8		12		25	3	45	
7	1	11	1	20	1	40	
6	1	10		15	2	35	
5		9		10	3	30	1
4	1	8		5	9	25	3
3	3	7	1	0	4	20	3
2	5	6	1			15	3
1	13	5	1			10	7
		4	2			5	5
		3	2				
		2	5				
		1	5				
		0	5				

one and the total time spent on the other two is too slight to be of service in estimating the validity of the tests, nor can the sum of the three be regarded as a serviceable measure of persistence in general.

The superintendent of the orphanage rated the children for us as to their self-control in the ordinary situations of school life on a scale of 1 to 5. The separate persistence tests show a slight disagreement with these ratings, suggesting that the persistence tests are not measures of self-control. The three together correlate − .57 with this criterion.* That is, those who persisted longer at the three assigned tasks had lower self-control ratings, rather than higher, as might be expected.

* The *rho*'s with the ratings (ranked) are, respectively, Balls and Pennies, − .33; standing on one foot, − .30; and working on puzzle, − .27. The correlation of − .57 reported for the sum is the *rho* transmuted into *r*.

Tests of Inhibition. There were seven of these tests, including the pain balance task. The latter was scored, as will be recalled, in terms of the difference between the weight recognized as uncomfortable, called the "pain limit," and the weight regarded as the limit of endurance. The average amount of weight at the endurance limits was 2860 grams. By inspection of the actual differences between the discomfort and pain limits, a series of differences varying in amount with the size of the discomfort limit was arbitrarily taken as indicating the presence or absence of the power to endure pain; and these "fact" scores were added to the "yes or no" scores on the other inhibition tests in getting a "fact" total consisting of the number of tasks the individual passed successfully. "Passing" meant going through the experiment without any change in facial expression. Change of expression meant failure to inhibit. As there were seven tests, these scores ranged from 0 to 7.

A total amount score was also assigned based on the scaled values given to each reaction, ranging from 0 to 3, as already described. The pain weight task was not included in these totals. Consequently there was left a total range of 0 to 18.*

The separate tests of inhibition were considered in relation to one another and to the total score. Table LI shows their relation to one another. The intercorrelations range from − .094 to + .775, averaging + .225. The relative size of the correlations indicates the amount of association between the various behaviors. Those that seem to hang most closely together are the responses to funny pictures and to bad odors or tastes. Next come responses to funny pictures and to tickling, and to tickling and smell. Taste and smell also are associated to the extent of .429. The responses to the startler and candy, however, are not associated appreciably with any of the others.†

* These two methods of summing gave a rank difference *rho* of .94.

† The average inter-*r* of .225, when substituted in Formula (12), page 120, Book Two, *Studies in Deceit, viz.*,

$$\text{Validity} = \sqrt{\frac{Tr}{1 + (T - 1)r}},$$

yields a predicted validity for these six tests of .78. Omitting the Candy and

TABLE LI

INTERCORRELATIONS OF INDIVIDUAL INHIBITION TESTS

TEST	2	3	4	5	6
1. Feather	− .094	.206	.490	.490	− .136
2. Startler		.165	− .045	.145	.042
3. Taste			.429	.775	.089
4. Smell				.632	.041
5. Humor					.041
6. Candy					

The extent to which the response on each of the six tests reported in Table LI is an indication of response on all of them together is shown in Table LII.

TABLE LII

VALIDITY OF INHIBITION TESTS AS MEASURED BY THE SUM OF SIX

TEST	r*
1. Feather	.608
2. Startler	.406
3. Taste	.775
4. Smell	.804
5. Humor	.842
6. Candy	.506

* These are rank difference correlations in the form of the corresponding r's. But as the criterion contains the test with which it is correlated, they are somewhat too high, especially those for Smell and Humor.

If the six together can be regarded as a measure of inhibition, this table shows the validity of each of the six tests taken one at a time. The response to the funny pictures evidently stands out well above the rest, with that to smell a close second, then to taste, and fourth, to tickling with the feather. This last correlation of .608, of

Startler tests, the average of the other four is .50 and their theoretical validity is .89.

course, is based on too small a number for generalization (it has a probable error of ± .09); but in any case it is too low to enable one to predict whether a child who could keep a straight face when tickled would also similarly inhibit other strong tendencies in proportion.

Our final source of information about the tests is in the ratings by the superintendent, to which we have once before referred. The correlation of the entire battery of seven tests with this criterion of the power of self-control is .55. The separate tests correlate with this criterion to the extent shown in Table LIII.

TABLE LIII

VALIDITY OF INHIBITION TESTS AS MEASURED BY THE SELF-CONTROL RATINGS

TEST	r*	r†
1. Feather	.126	.125
2. Startler	− .261	− .52
3. Taste	.426	.70
4. Smell	.063	.33
5. Humor	.188	.40
6. Candy	− .188	.23

* These coefficients are *rho*'s transmuted into *r*'s. The scores showing degree of reaction are used.

† These are biserial *r*'s, counting any reaction as failure to pass the test.

The two tests that are negatively associated with this criterion are the same two that give the lowest correlations with the sum of the six and which show occasional negative intercorrelations with the other tests.

These data, although based on only twenty-five cases, are very suggestive of the possibilities of individual testing in this field and offer a most promising research problem.

CHAPTER XXII

TESTS OF INHIBITION IN PARTY GAMES

THE GENERAL PLAN

In setting up test situations for measuring deception, we found that a party could be planned so as to turn some of the usual stunts into deception tests. This suggested to us the possibility of using parties for measuring inhibition also. Many of the stunts that children enjoy doing on such an occasion find their interest in the fact that some people cannot control themselves in certain situations. The old game of trying to cause someone else to laugh by making faces at him is a good illustration, or "Simon says, 'Thumbs up.'" To be entirely satisfactory for test purposes, such behavior should be susceptible of quantitative measurement; but our good luck with the individual tests, described in the last chapter, which depended on ratings rather than records, made us feel that it would be worth while to try out the Inhibition Party idea.

With the aid of a graduate student and one of our staff* we developed a standard party to last about one hour. During this time, every pupil was subjected to the following tests:

1. The Hopping Race
2. Crows and Cranes
3. The Funny Story Game
4. Simon says, "Arms Up"
5. The Noise Response Game
6. The Hall Test (used only on a few groups)

---

* Captain C. W. Christenberry and Chester A. Kerr.

The complete directions for administering the parties are printed in Appendix H, but we shall describe each test here and explain the general plan.

The parties were given in school C, where there is a large gymnasium on the top floor. This was not closed off from the hallways, and pupils changing classes had to pass through. This required special arrangements that would not be needed elsewhere. As far as possible the regular classroom gymnasium period was utilized, extended to one hour. As this time was occasionally given over to a social event, the use of it for the games and stunts was a perfectly natural occurrence from the children's standpoint.

As each class came up to the gymnasium, the game leader said:

> Your principal has agreed that it will be a nice thing for us to show our appreciation to those pupils who have been taking our tests by giving them a party here in the gymnasium. During this next hour your room are guests at a party. There are to be games that will give each of you the chance to win a real prize. There is one thing, however, that we must ask of all of you before we go on with this plan. As we have only one hour, it will be impossible to carry out the plans for the party unless each of you is willing to do just as you are told and follow all instructions carefully. Will all who understand this please raise one hand? Remember, do just as you are told, and the party will be a pleasant time for everybody.

The class was then divided into two groups, the Reds and the Yellows, and each child had a badge with a red or yellow number, by which he could be identified during the games. The test games were given to one color at a time while the other half of the class were entertained by one of the leaders in a group game until their turn came.

There were enough leaders or judges for each party so that in watching the children each judge would need to record the behavior of not more than six subjects. Each judge had a slip of paper for each stunt, with the numbers of the six children whom he was to observe. His only responsibility was to watch for failures to inhibit, as will be described presently, and make a check mark oppo-

site the number of the subject in the column representing the trial on which any such failure was made.

DESCRIPTION OF THE TESTS

The Hopping Race. The point of this stunt is to see which subjects can wait for the signal to start when the expected signal is delayed. The pupils are lined up back of a chalk mark and told to start and hop on one foot to the goal line a few feet away. In three heats, they start on the count " three." In the fourth heat, the instructions are to start on the word " six," instead of " three," and this is repeated once, making five trials in all. Of course the actual winner is kept track of as well as those who start before the right signal.

Crows and Cranes. The Yellows or Reds, as the case may be, are again divided into two groups, one called " Crows " and the other " Cranes." These line up about ten feet apart in two rows, facing each other. If the signal given is " Crows," the Crows try to catch the Cranes before the latter can get back to their goal line, about twenty feet away. Likewise, if the signal is " Cranes," the Cranes chase the Crows. The child who starts the wrong way on the signal is scored a failure.

The signal was given in such a way as to make prediction impossible. Thus: " All ready. On the lines. Cr-r-r-r-r-rows ! " Six trials were given, so that each pupil had six opportunities to make the wrong move.

The Funny Story Game. Both the Yellows and Reds combined for this stunt. The children formed a large circle, sitting on the floor. It was explained that the prize would go to the pupil who succeeded in keeping a straight face while three jokes or stories were told. Signals were given, *viz.*, " Hold " and " Let go," to define the limits of each test period. The judges watched certain subjects, as usual, and checked two levels of response — a smile and a laugh.

The difficulty with this test is the choice of the story. It is hard to predict what will seem funny to a particular group. A story

that arouses instantaneous response in one group will fall flat with another.

Simon Says, "Arms Up." This is a modification of the "Thumbs up" game, only arms are used as they can more easily be seen by the judges. The same signals are used as in the old game. The contestants are to move their arms up or down only on the complete command: "Simon says, 'Arms up,'" or "Simon says, 'Arms down.'" If a pupil moves his arms when the wrong command is given, as "Arms up," without the words "Simon says," he is checked as failing to inhibit.

The Noise Response Game. The whole class again formed a circle, standing face out. A variety of strange or startling noises was made, with a sharp whistle frequently interspersed. The pupils were to show no sign of having heard anything except on the whistle, when they were to turn around and face in. The first one to turn around when the whistle blew was the winner. Thus all were on the qui vive to respond. The other sounds used were: a bursting paper sack, a siren, a loud yell for help, two boards clapped together, the siren again, breaking some bottles. Irregular time intervals were allowed, with the whistle coming frequently.

The Hall Test. Bonbon souvenirs of the party were placed at the end of one of the school hallways. These old-fashioned gifts can be snapped to make a cracking noise. The pupils were asked to go quietly so as not to disturb other classes, get their souvenirs, and return to the gymnasium for the awarding of the prizes. They were kept under observation of course. As there were no exceptions to perfect order, this test was dropped after a few trials.

RELIABILITY OF THE TESTS

In each of the five tests that were administered to all groups tested in school C, five or six trials were given on each test. A score was recorded for each trial: 0 meaning perfect inhibition; 1, partial inhibition; and 2, complete failure to inhibit. In each case the score represents the judgment of the observer.

The reliability of each test was determined by correlating the odd-numbered trials against the even-numbered trials. In making these determinations we combined the scores "1" and "2" and used only two scores — calling all failures to inhibit 1 and successes 0. The average reliability of the five tests is .501.

TABLE LIV

RELIABILITIES OF THE PARTY INHIBITION TESTS

TEST	r ODD AND EVEN TRIALS *	PREDICTED RELIABILITY
Funny Story game	.578	.732
Noise Response game	.364	.533
Hop race	.341	.508
Crows and Cranes	.302	.460
"Simon says"	.160	.276
Average	.349	.502

* These correlations were all computed by a fourfold table. This method was used because the preponderance of zero scores so skewed the distributions as to render other methods unsafe.

In order to find the reliability of the whole party, we pooled all five games. There were 28 trials. The correlation between the 14 odd trials and the 14 even trials is .564, which gives a predicted reliability coefficient of .721 for 28 trials. These figures mean that fourteen trials spread over five test games correlate .564 with fourteen more trials — or that 28 trials (a whole party as we gave it) would correlate .721 with another party. In order to get a reliability of .90, we should need a party with a total of 72 trials in 12 games. Such a party obviously could not be run off in one hour. Since it would require two parties to secure sufficiently reliable results and since these parties were very expensive to administer, we were forced to abandon this particular feature of our testing program.†

† The low reliability of the party games procedure for measuring inhibition is doubtless due in part to the fact that the scores are records of *observation*. There are probably wide individual differences in observers. Some may be more severe than others in scoring, thus introducing a variable factor. Also the standard as to what constitutes failure to inhibit may shift somewhat from game to game. Hence part of the unreliability of these scores is due to errors of observation.

VALIDITY OF THE TESTS

Since we have no outside criterion of inhibition for this population, we can determine only the theoretical validity of the party procedure as a whole by intercorrelating the separate tests. In this sense, the party had practically no validity. Although the results of twelve games such as we planned would agree fairly well with the results of the same games repeated, they would not agree at all well with the results of any number of *similar* games. Indeed the average intercorrelation of the five games we used is .03. This indicates that these five party tests each measure an almost totally different kind of thing. Evidently waiting for the signal in the hop race reveals a different kind of inhibition from starting on the right signal in the Crows and Cranes game, or obeying the commands in the "Simon says" game, or repressing laughter, or inhibiting the tendency to turn the eyes in the direction of an unusual noise — all these are almost totally uncorrelated in the ten groups tested.*

RESULTS OF THE PARTY TESTS

There were, of course, wide classroom differences in the way the children responded to the test situations. It is not necessary to go into these differences here, as our purpose in administering

* These zero inter-*r*'s are very difficult to interpret. They are not wholly due to the unreliability of the tests, for when corrected for attenuation none rises above .25. Were it not for the fact that each game has a certain degree (some a high degree) of reliability, one might surmise that we were only correlating errors. But we are not. We must conclude that this type of behavior is highly specific. The complete set of inter-*r*'s is given in Table LV.

TABLE LV

INTER-*r*'S OF PARTY INHIBITION TESTS

TEST	2	3	4	5
1. Hop Race	+ .017	+ .049	+ .018	− .047
2. Crows and Cranes		− .166	+ .053	+ .051
3. Funny Story game			+ .081	+ .073
4. "Simon says"				+ .175
5. Noise game				

these parties was only to give this technique a preliminary trial in order to decide upon its usefulness for our more extensive program in populations X, Y, and Z. Certain features may be of interest, however, to anyone contemplating the adoption and improvement of this method of testing. We shall report two sets of facts only. The first is the per cent of the 310 pupils in grades 4 to 8 who failed to inhibit on each trial of each test. The second is the percentage distribution of inhibition scores for each test.

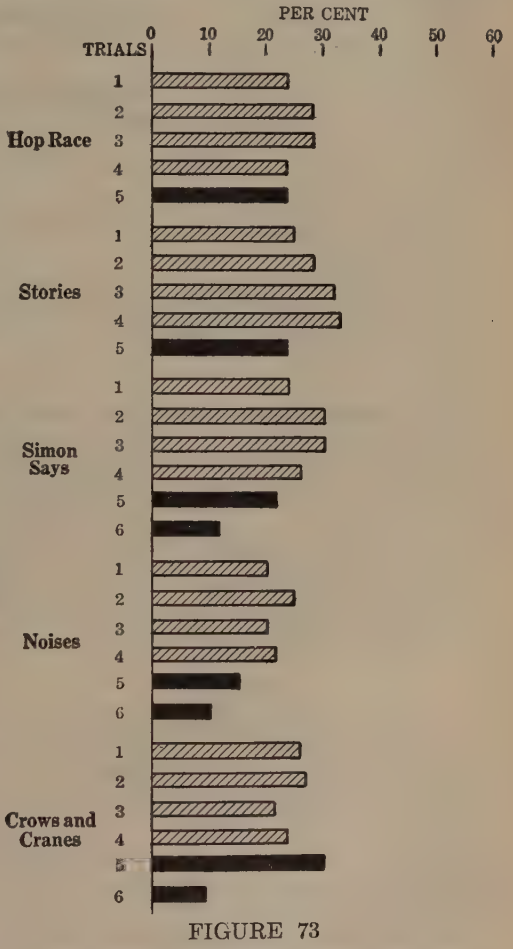


FIGURE 73

PER CENT FAILING TO INHIBIT ON EACH TRIAL OF EACH PARTY TEST

Per Cent of Failures. These facts are portrayed in Figure 73. From this it stands out plainly that the last trial showed fewest failures and that the most failures occurred somewhere in the middle of the test. Further study of this phenomenon might reveal a general tendency toward waves or rhythms of attention.

Percentage Distribution of Scores. These distributions

are much alike and may be represented by the upper histogram in Figure 74. Evidently the tests were too "easy" — that is, too large a proportion of the children could successfully resist the tendency to laugh or move, no matter what stimulus was used. What is needed is a much more stimulating temptation, such as actual tickling, or an itch powder, or snuff, or a physically painful sensation; but this was obviously out of the question. When the tests are all combined, however, as in the lower histogram of Figure 74, the distribution is more nearly normal. It rises in the middle and tapers off on each side. This is possible because of the low relation among the tests — a child who inhibits well on one test may yield on the rest.*

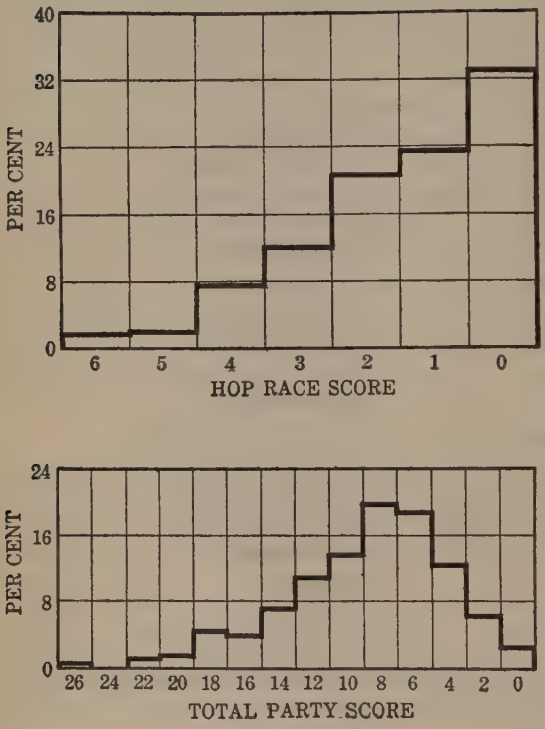


FIGURE 74

PERCENTAGE DISTRIBUTIONS OF HOP RACE SCORES (UPPER) AND TOTAL PARTY INHIBITION SCORES (LOWER)

Decreasing scores represent increasing inhibition.

* In view of the large amount of error in these party scores, this normality of the distribution of the total score might be regarded as a function of the errors rather than of the scores, the errors being uncorrelated.

PART II

Factors Associated with Self-Control

CHAPTER XXIII

AGE, SCHOOL STATUS, AND INTELLIGENCE

The general plan of the Character Education Inquiry calls for sample studies or exploratory investigations over a rather wide area of interest. The largest of such units of work is our *Studies in Deceit.* In Book One of this volume we presented the results of a somewhat less extensive study of service. Book Two, although presenting a fairly extensive array of self-control tests — some twenty in all — devotes even less attention than Book One to the consideration of the relation of the behavior in question to other facts. It will be possible for us to condense our material at many points on account of the similarity of the treatment to that of Book One, which is already familiar to the reader. The order in which the various factors dealt with are taken up conforms to the order followed in Book One.

AGE AND SELF-CONTROL

At the outset it should be made clear that it is not our purpose to establish age norms for any of these tests from the data available. At best all that can be done is to find out whether or not the older children tend on the whole to score higher or lower than the younger children. If it turns out that the older groups score higher than the younger, other things being equal, then the ground has been cleared for a more exhaustive age study on many more cases, which might result in establishing age norms.

In presenting the available data we recognize the fact that "chronological age" is a complex affair. It is mixed up with school status, with mental age, with social experience of all sorts.

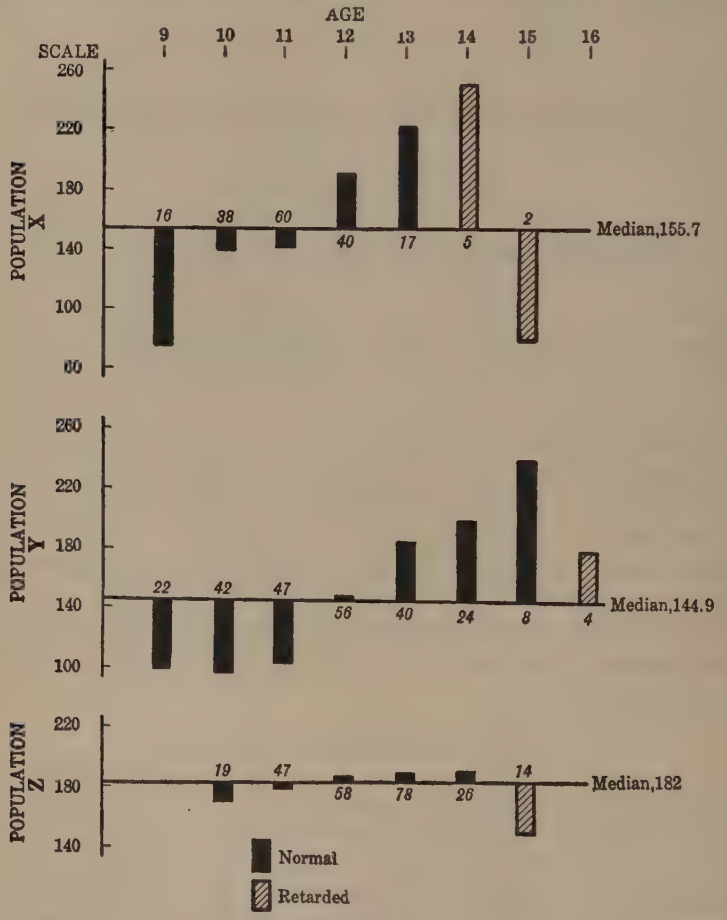


FIGURE 75

MEDIAN TOTAL PERSISTENCE SCORES BY AGE COMPARED WITH THE MEDIAN FOR ALL — POPULATIONS X, Y, AND Z

The N for each group is shown in small figures at each bar.

It is commonly assumed that " older " means not only biologically more complex but also " more experienced."

While it may be safe to assume that biological growth is on the whole favorable to growth in character, the particular conflicts found in any civilization between cravings and social conventions are necessarily functions of growth and may increase rather than decrease, thus introducing a biological factor not positively associated with age.

Whether " experience " will tend to increase or decrease scores on character tests depends on what kind of experience it is. In case of scores on intelligence tests almost any kind of added school experience will increase them. But with character tests the case is different. Of two children with a certain type of experience the older has an advantage, whereas a second child of equal age but with other types of experience is at a disadvantage. Of all the possible kinds of experience there are probably only a few that will actually decrease intelligence, but there are probably many that will pull down character.

Age and Persistence. Figure 75, which is comparable to Figure 14, page 122, shows the median total persistence score for each age in comparison with the median for the total population. As may be seen from the table in the accompanying footnote,* the tests

* The regression lines are linear; therefore the r's of Table LVI may be taken at their face value.

TABLE LVI

CORRELATIONS BETWEEN PERSISTENCE TESTS AND AGE — POPULATIONS X, Y, AND Z

TEST	POPULATION X	POPULATION Y	POPULATION Z	AVERAGE
Story Resistance . .	− .003	.280	.084	.12
Cross and Ring . . .	.402	.331	.181	.30
Magic Square . . .	.158	.391	− .216	.11
Persistence for Self .	.208	.119	.060	.13
Persistence for Class .	.178	.104	− .195	.03
Total persistence . .	.391	.438	.169	.33

Correction for restricted range would be unwarranted without further

vary among themselves in the extent to which the scores are associated with age. We have considered certain possible explanations of the differences among the schools and the differences among the tests, but without arriving at any sure conclusions. The general drift is clearly toward increase with age, and age norms for certain specific sorts of persistence can be calculated at any time by testing large enough numbers of children.

Population Differences in Correlations between Age and Persistence. In the X and Y populations we may be certain that the older children are the more persistent, but the case of Z is doubtful. In two of the five tests the *r*'s are negative, and in the other three they are low positive. It might be thought that increased experience in this underprivileged social environment does not tend to increased persistence of the types measured by these tests. But when the tests are combined, the medians for Z do increase steadily with age, though by no means as rapidly as in the case of X and Y, where the increases are very marked. In X and Y the medians jump 100 points, or over half the scale, from the ten-year-old group to the fourteen-year-old group. In Z the jump is only 20 points. But in Z the ten-year-olds are already at a median of 170, while in X they are at 140 and in Y 98.

The fact that the whole population Z is higher in persistence than either X or Y suggests exactly the reverse of the influence just suggested, that this type of environment does not favor persistence, for it seems to favor it very much. In fact it favors it so much that by the time the Z children are ten years old they score as high as the twelve-year-olds of X and the fourteen-year-olds of Y. That is, measured by X and Y, they are from two to three years accelerated. This may be one reason why they do not appear to increase so much from ten up, but of this we shall speak more fully later.

Again, however, caution is necessary, for we find that Z's superiority among ten-year-olds is only on the Puzzle tests and the Persistence for Class test. We know from previous investigation that the fifth-grade class in Z was for some reason extremely coöperative, which may account for these high scores in Persistence for Class. Furthermore, Table LXXI of Appen-

---

knowledge concerning the linearity of the regressions beyond the age limits included. Correction for heterogeneity of class groups would be desirable, as classroom differences are sometimes considerable.

dix C shows how low their score (grade 5) was on the Persistence for Self test. As for persistence on the Puzzles, it might be thought that, being of only moderate intelligence, they were not interested in the puzzles but dawdled along in a faint-hearted fashion without knowing they were beaten.*

Age and Inhibition. The relation between age and inhibition is slight and variable. When all the inhibition tests are combined in the manner described in Chapter XVII, we find r's of .015, .134, and .136 between age and inhibition for populations X, Y, and Z. For boys alone, all the r's are zero. For girls alone they run .091, .322, and .220. Here the increase of the ratio of "reason" to "desire" as age increases is apparent in Y and Z. The medians for all age groups are printed in Table LXVI of Appendix C.

As we know the correlation between persistence and inhibition,† we can partial out the factor of increasing interest (so far as this is represented by the persistence scores) from the correlation between inhibition and age to find what the association between age and inhibition would be if all ages were equally persistent. This r is $-$ .022 for X, $-$ .037 for Y, and .108 for Z.

Age and the Separate Inhibition Tests. The Picture test by itself shows a small but consistent correlation with age: .137 for X, .175 for Y, and .127 for Z. But this test involves speed of adding, and we know that this ability correlates to a far greater extent than this with age, so that, if the factor of initial ability is taken into consideration, the inhibition feature by itself is seen to be unrelated to age.‡ In the case of the other tests, all the r's are prac-

* This may be true of the older groups as well, though to a lesser extent; but if this interpretation should prove correct, it would help to account for Z's high persistence means at certain ages and would show them to be, in part at least, fallacious. The weakness of these two tests under these conditions would be obvious. The indications, however, are against this interpretation. In population Z, for example, the accelerated children were the most persistent and the retarded the least so.

† The r between persistence and inhibition totals is .091 for X, .378 for Y, and .191 for Z.

‡ Sustained-speed ability also correlates from .25 to .65 with the Picture Inhibition test — a matter to be discussed later. The facts might be inter-

tically zero except for the Story Inhibition test in population Y. Here we find the older children resisting the temptation to open the sheets more frequently than the younger children. But the question is whether the older children were as keenly interested in the stories as the younger ones. The Picture and Story temptations run counter to school drives. It is possible that the older children have greater interest in getting good grades than the younger. In population X, where motivation might be regarded

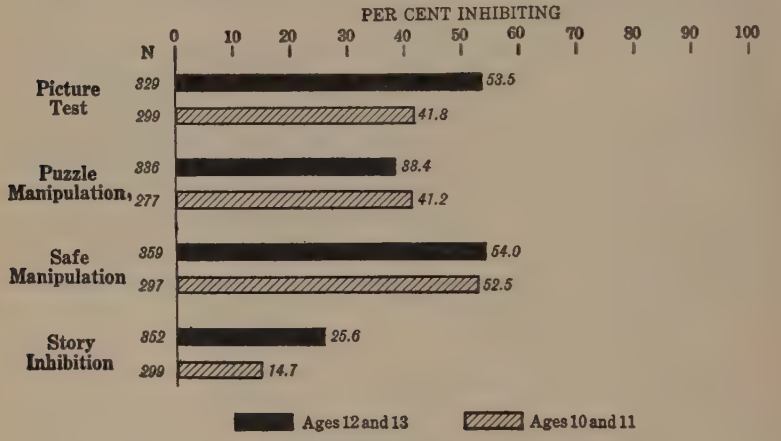


FIGURE 76

COMPARATIVE FREQUENCY OF INHIBITION AMONG CHILDREN TEN TO ELEVEN AND CHILDREN TWELVE TO THIRTEEN POPULATION XYZ

as better distributed through the grades, there is no relation whatever between age and the scores on the Story Inhibition test.

Figure 76 compares all pupils ten and eleven years old with all pupils of twelve and thirteen in the percentage successfully over-

---

preted to mean, therefore, that the ability to keep up a rapid rate in adding is itself partly a function of the ability to *resist distractions* in adding. The speed aspect increases with age, but not the power to inhibit distractions. With speed ability constant, the *r*'s between the Picture test and age drop to .034 for X, − .052 for Y, and .042 for Z.

coming the temptations presented by each test. A substantially larger * proportion of the older children than of the younger children succeed in resisting the tendency to look at pictures, handle puzzles, or read stories when the alternative is monotonous rapid additions or letter counting for the sake of a good record.

GRADE STATUS AND SELF-CONTROL

Each school grade contains a wide range of ages, but the ability to gather the enlarging experience implied by growth in years is perhaps better indicated by a child's grade than by his age. As we have already stated, we recognized the necessity for distinguishing between those who are progressing normally from grade to grade and those who are for any reason retarded or accelerated. Also there is undoubtedly a fundamental difference in social adjustment represented in scholastic and deportment marks which needs to be taken into consideration in any study of school status. With these several factors we shall now deal.

School Grade and Self-Control. As might be expected from the correlations between age and persistence, we find large grade differences in the total persistence scores. This is especially true of X and Y. The facts are shown in Figure 77, which is comparable to Figure 75. The complete tables are printed in Appendix C (LXVII–LXXII).

In the case of inhibition the grade differences shown in Figure 78 are greater than the correlation with age would lead us to expect. (See also Tables LXXIII and LXXIV, Appendix C.) Advancement in school is apparently associated quite definitely with increasing tendency to complete a task once begun and with increasing power to inhibit passing interests and distractions in favor of less personal or immediate satisfactions.†

---

* The difference between the percentages in the Picture test is 3 times its SE; and in the Story Inhibition test, it is 3.5 times its SE.

† For inhibition the difference between the lower and upper grades is 4.1 σ. For persistence it is 10.1 σ.

Retardation and Acceleration. The facts so far reported have paid little attention to the problems of adjustment arising from poor grading. In discussing the service tendencies of children we reminded the reader of the common observation that children who are atypical in some respect are likely to offer more behavior problems than those who are more normal, and we adduced as confirma-

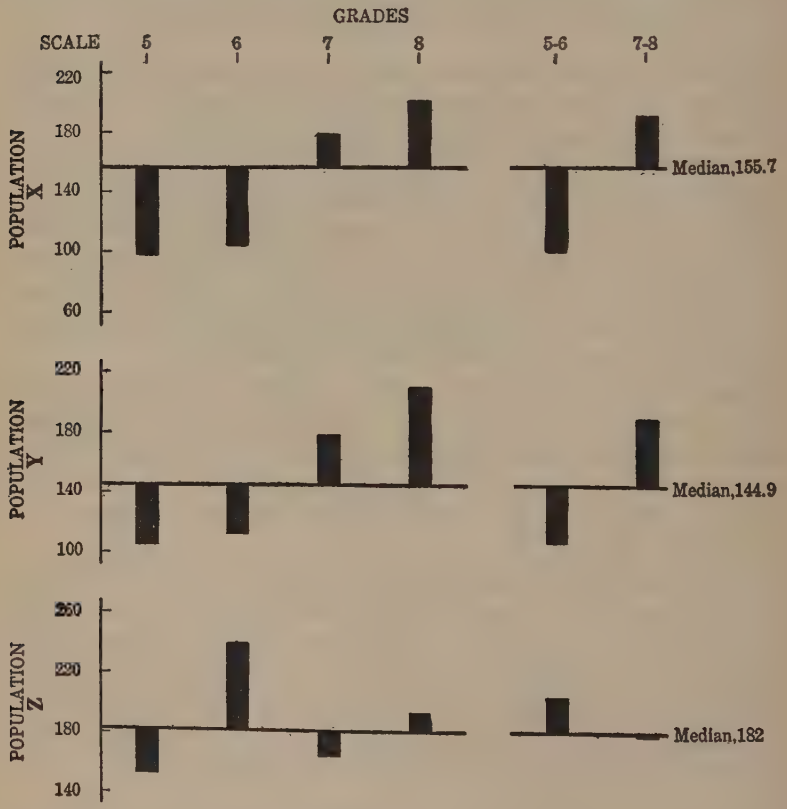


FIGURE 77

MEDIAN TOTAL PERSISTENCE SCORES BY GRADE COMPARED WITH THE MEDIAN FOR ALL — POPULATIONS X, Y, AND Z

The data for this figure are given in Table LXVII, Appendix C.

tory evidence the scores of retarded children on the service tests. But we found the accelerated children on the whole slightly *more* coöperative and helpful than those who were near the average in their school progress.

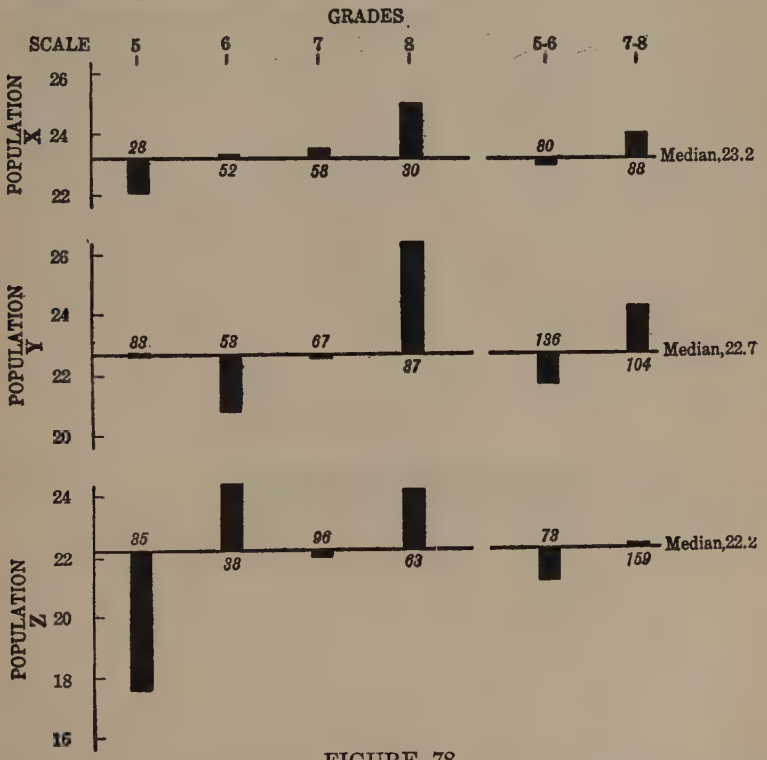


FIGURE 78

MEDIAN TOTAL INHIBITION SCORES BY GRADE COMPARED WITH THE MEDIAN FOR ALL — POPULATIONS X, Y, AND Z

The N for each group is shown in small figures at each bar.

In our study of self-control we shall deal with both scholastic grade status and social grade status.

Scholastic Grade Status and Persistence. In attempting to discover the facts regarding retardation and acceleration in relation to

persistence, we made distributions of the total persistence scores for each age level in each grade. We then computed the median score for each age group. This gave us the figures in Table LXXV of Appendix C. In order to even up the grade differences, we then made a rough classification of pupils into three groups, normal, accelerated, and retarded, according to the general appearance of the distributions, and then combined all the normal, all the retarded, and all the accelerated, irrespective of grade.

We find in general that the retarded are the most persistent, partly or entirely because they are in general older than their classmates. In population X, the accelerated, normal, and retarded groups have median persistence scores of 130, 156, and 180, respectively. In Y the scores are 124, 140, and 157. In Z, however, the order is 189, 174, and 167. For all combined it is 148, 158, 164.

Social Grade Status and Persistence. Following the method described on page 134, we put in one group all the children of

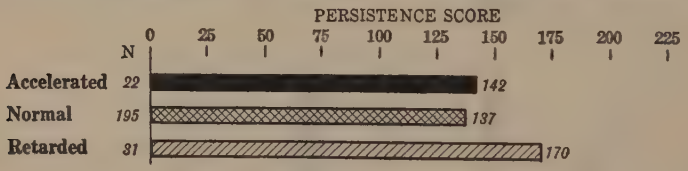


FIGURE 79

MEAN PERSISTENCE SCORES OF ACCELERATED, NORMAL, AND RETARDED PUPILS OF POPULATION Y

population Y who were over age for their class and in another group all who were under age, keeping boys and girls separate. The results are shown in Figure 79. In this population the children of both sexes who are retarded in comparison with their own classmates are more persistent than the normal and the accelerated, but the latter are slightly superior to the normal. These differences, however, are not statistically substantial.*

In view of the tendency in Y for the retarded to be somewhat more persistent than the normal and the accelerated, it might be

* The difference between the accelerated and retarded is 1.3 times the SE of the difference.

thought either that our correlations between age and persistence are really functions of retardation or that the superior persistence of the retarded is due simply to the fact that these children are older than their classmates. That there is a mutual relation between these facts is probably true, but the dominant factor seems to be the association between persistence and age. This would be indicated by the fact that, when the accelerated and retarded are omitted from the scattergram for Y, the correlation between age and persistence is .475, which is higher than when the accelerated and retarded groups are left in. Furthermore, in population Z, although the order in persistence is accelerated, normal, retarded, the correlation between persistence and age is positive, though much lower than in X or Y (.169). All this would seem to indicate that the superiority of the retarded over the normal and accelerated, in the case of X and Y anyway, is chiefly a function of age.

It would be desirable to compare the intelligence scores of these three groups to see whether the accelerated are enough more intelligent than the others to warrant a higher score than they have, since intelligence and persistence are correlated. As we shall presently show, however, the correlations of persistence with intelligence are much lower than with age, so that we should not be likely to find that acceleration and retardation in and of themselves play more than a minor rôle in determining general persistence.

Scholastic Grade Status and Inhibition. The differences are small and favor sometimes the retarded, sometimes the accelerated. The net result is approximately zero. The facts are given in Table LXXVII, Appendix C.

Social Grade Status and Inhibition. The facts are shown in Figure 80. The accelerated girls and the retarded boys show greatest power of inhibition. If the sexes are combined, accelerates, normals, and retardates are indistinguishable.

INTELLIGENCE AND SELF-CONTROL

Intelligence and Persistence. Using a measure which has the same significance as an IQ, we find a positive correlation between

intelligence and persistence. The r's for the three populations are: for X, .244; for Y, .226; and for Z, .091. For all combined the coefficient is .147. Reputation for persistence is still more highly correlated with intelligence, the r's running .402, .214, .430, and .356. For the separate tests, the largest r's with intelligence are for the work done on the Stories and for the tendency to keep up the monotonous additions when working for the class. Intelligence might readily be a factor in the interest shown in the Stories, but hardly in the case of the Persistence for Class test.

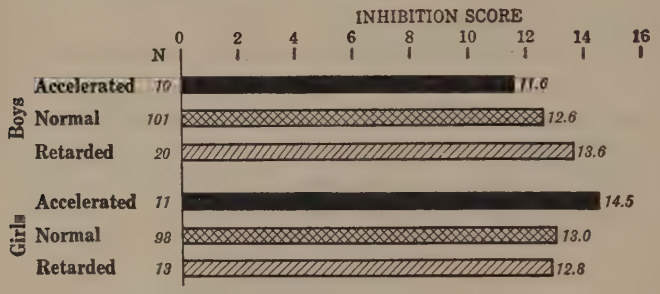


FIGURE 80

MEAN INHIBITION SCORES ON THREE TESTS OF ACCELERATED, NORMAL, AND RETARDED BOYS AND GIRLS OF POPULATION Y

Mixed up with the association between intelligence and persistence is the factor of age, which we have shown to be positively correlated with persistence. Our intelligence scores are similar to intelligence quotients, and in the type of population represented in any fixed group of public school children, this type of intelligence score usually decreases slightly with age, on account of the tendency for the older pupils to be those who are duller and therefore left behind. We have already suggested that the higher persistence scores for the older children may largely account for the correlation between retardation and persistence. Hence, if we keep age constant, we shall have a far more accurate picture of the true relation between intelligence and persistence. The partial r, showing the amount of association between intelligence and persistence among

children of the same chronological age, is .366 for X, .306 for Y, and .126 for Z.

Figure 81 summarizes the relations between intelligence and persistence for population Y and should be compared with Figure 79. Table LVII in the accompanying footnote gives all the relevant r's.* In Figure 81, the length of the bars shows the median per-

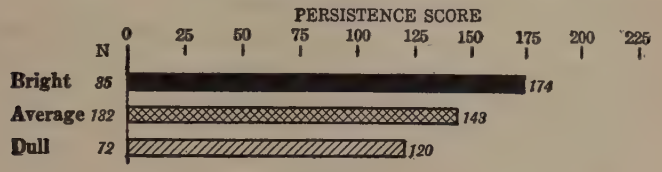


FIGURE 81

MEDIAN PERSISTENCE SCORES OF BRIGHT, AVERAGE, AND DULL PUPILS OF POPULATION Y

sistence score for the level of intelligence indicated at the left. For convenience, three bars are used, representing the brightest, the middle, and the dullest. The tendency for persistence scores to increase with intelligence thus becomes clear, by contrast with Figure 79, which shows the retarded as most persistent.

*TABLE LVII

CORRELATIONS BETWEEN INTELLIGENCE (CAVI SIGMA) SCORES AND PERSISTENCE SCORES — POPULATIONS X, Y, AND Z

TEST	POPULATION			
	X	Y	Z	Total
Stories	.223	.135	.258	.176
Cross and Ring	.111	.232	.020	.047
Magic Square	.218	.281	− .074	.077
Persistence for Class	.133	.082	.161	.123
Persistence for Self	.044	− .026	.060	.044
Total persistence	.244	.226	.091	.147
Reputation	.402	.214	.430	.356

Intelligence and Inhibition. Intelligence is not closely associated with our inhibition scores, save in population Z. Here the association is striking for the Puzzle and Safe Manipulation tests, as may be seen by consulting Table LVIII in the accompanying footnote.* The *r*'s for the total inhibition score are .071 for X, .003 for Y, and .269 for Z. Reputation for inhibition is a little more closely correlated with intelligence.

Thence we see that intelligence in the sense of ability for one's own age is positively associated with persistence but is only slightly related to inhibition except in population Z. The children of greater intellectual ability have some advantage when it comes to the effort to overcome obstacles or to apply themselves to a monotonous task, whereas in the choice of the social rather than the self-centered behavior required by the inhibition tests the children of less intelligence have no handicap. Of course, this does not mean that the brighter pupils could not *learn* habits of self-control more readily than dull pupils. It means only that it is not apparent that they have done so. It may be fairly assumed that efforts to teach or inculcate these habits have not been undertaken in the populations tested.

* TABLE LVIII

CORRELATIONS BETWEEN INTELLIGENCE (CAVI SIGMA) SCORES AND INHIBITION SCORES — POPULATIONS X, Y, AND Z

TEST	POPULATION		
	X	Y	Z
1. Puzzle Manipulation †	.000	− .089	.306
2. Safe Manipulation †	.162	.134	.349
3. Story Inhibition †	.169	.200	.000
1, 2, and 3 combined	.066	.040	.374
4. Pictures	− .047	− .031	− .046
1, 2, 3, and 4 combined	.071	.003	.269
Reputation for inhibition	.281	.168	.320

† Biserial *r*'s.

TEACHERS' MARKS AND PUPILS' SELF-CONTROL

One might readily guess that good grades are the result of intelligence, hard work, and the ability to resist distracting stimuli — and of good standing with the teacher. We have measures of all but the last and can see whether the types of conduct included in our tests conform to this expectation.

Teachers' Marks and Persistence. The extent of the tendency for more persistent behavior to be associated with higher school grades is shown by the *r*'s of .204 for Z, .118 for Y, and actually − .004 for X. It would seem that hard work of the sort represented on the tests is more apt to get results in marks in a population where the general level of intelligence is low (Z) than where the general level is high (X). Being somewhat more alike in capacity, the pupils of Z become differentiated in achievement by virtue of the tendency to overcome obstacles.

The absence of relationship between persistence and school marks in population X is interesting, as it corresponds to investigations made of amount of time spent on study as compared with grades received.* The X pupils constitute more nearly the type found in college than do those of either Z or Y.

But the teacher's opinion of the child's school achievement is expressed also in deportment marks, and for our study these are of perhaps greater importance than the academic grades. The industrious child surely ought to be accorded a higher position on a scale of school virtues than one who does not trouble to complete his tasks. As a matter of fact, however, the *r*'s between deportment and our persistence scores are not high: .155 for X, .263 for Y, and .149 for Z.

Teachers' Marks and Inhibition. Resistance to distracting stimuli makes little difference in academic success in X and Y, but is of marked significance in Z. The *r*'s are − .033 for X, .112 for

* In an unpublished study of the time spent by theological students on class work the authors found the *r* between study and marks to be zero. May reports the same facts for university students in "Predicting Academic Success," *Journal of Educational Psychology*, 1923, **14**, 429–440.

Y, and .371 for Z. For the Story Inhibition, Puzzles, and Safes, alone, the r for Z is .431. The median total inhibition score of the 89 cases (19 of whom were in Z) marked 90 to 99 in school work by the teachers of all three populations is 24, whereas the median inhibition score of the 39 pupils (30 of whom were in Z) marked 50 to 64 by the teachers is 17.3 — a difference which is one-seventh of the total scale.*

As for deportment, surely we might expect a correspondence between what the teachers would regard as good behavior and the ability to tend to business in the presence of attractive alternatives. And here the r's are considerably larger than for persistence, running for X .258, for Y .314, and for Z .306.

SUPPLEMENTARY DISCUSSION OF INTEREST AND SELF-CONTROL

Scores on our persistence tests increase with age (r's of .391, .438, and .169 for X, Y, and Z). These scores in each case represent a balance of interest. The interest in the solution of puzzles, in endings to stories, or in getting a prize for oneself or for one's class may vary with age in such a way as to make the older pupils appear more persistent when we ought more properly to say that they are more interested. If the interests appealed to could be varied sufficiently in character and as between the intrinsic appeal of the task and some end to be gained by doing it, much more could be learned concerning the nature of what we call persistence.

Similarly, our inhibition scores increase slightly with age in the case of the girls (zero correlations for boys, but .091, .322, and .220 for girls in X, Y, and Z); but again, it may be because the interests which need to be inhibited grow weaker with experience. Story endings, puzzles, and funny pictures may seem less alluring to the more sophisticated. If it is a poor rule that won't work both ways, this must be a poor rule, for we certainly cannot attribute the correlations with age in the persistence tests to *increasing* interest in puzzles and stories and then turn around and attribute the correlations with age on the inhibition tests to *decreasing* interest in puzzles and stories.

If we assume that persistence is a function of interest in the activity

* About 1.3 times the σ of the scale of inhibition scores.

concerned, then it would seem that the older children are better able to resist these interests when a sensible reason is given for doing so, for even though the interest in the activity increases, the power to resist it also increases.

If we assume that the increasing inhibition scores are due to falling interest in the activity, then the persistence correlations are doubly significant, for in spite of such weakening interests, the older children are more persistent.

Increasing experience as represented by grading is naturally also accompanied by increasing persistence scores. Interest would operate most strongly in the Puzzle Mastery tests, and it is here that the greatest grade differences occur. In the Persistence for Self and for Class tests, one would expect interest to increase less rapidly than in the case of puzzles, and as a matter of fact persistence here does not change much from grade to grade. Everyone can do something on these tests, but on the Ring and Cross test or the Story Resistance test, the younger children might feel that they could do nothing at all and so quit easily.

This phase of the matter may be checked by the correlations between mental age and the several tests. For the combination of the Cross and Ring and Magic Square tests these r's are .378, .515, and $-$ .014 for populations X, Y, and Z; and for the sum of the Persistence for Self and Class tests, they are .041, .059, and .089. For the Story test alone, the r's are respectively .371, .305, and .252. Mental age certainly makes a big difference in populations X and Y. In Z, however, those of lower mental age do not show any less inclination to work on the monotonous tests than on the mastery tests.

These facts may be interpreted in confirmation of the proposal that experience increases persistence because of increasing interest, since those whose ability is less mature have lower persistence scores on those tests which appeal to intellectual interests but keep up with the rest on tests making no special demands on intelligence; and of children of any one age it is the less able who drop out on the hard tests. When chronological age alone is considered, keeping mental age constant, the older pupils still show greater persistence than the younger, but not so much greater as when the effects of increasing age and increasing mental ability are combined.* On the whole, it seems to be the case that in population Y, and

* For Y alone, the r between chronological age and the Cross and Ring test is .331. With our CAVI point score constant, this falls to .202, which may be compared with the r between persistence and our CAVI sigma score

probably also in X, the persistence scores are functions of increasing chronological age, maturing mental ability, and native brightness in about equal degree.

CONCLUSIONS CONCERNING AGE, GRADE, AND INTELLIGENCE

The intermingling of the influences summed up in age, grade, and intelligence presents a complicated picture. In the case of persistence we find that the retarded pupils work longer at a given task than the rest, but this seems to be chiefly because they are older and age is definitely correlated with persistence. The r's of .391, .438, and .169 for X, Y, and Z indicate a genuine association between the sorts of conduct included in our persistence tests and chronological age. The correlation is greater with some tests than with others.

Intelligence is also associated with persistence (.224, .226, and .091, but .366, .306, and .126 with age constant). It is on the tests requiring intellectual effort that this association appears, however; so we may infer that the *interrelations* of age, intelligence, and persistence are specific, as well as the behaviors themselves.

The grade changes in persistence correspond to the observed

(partial, with age constant), which is .285. The r between the CAVI point score and the Cross and Ring test is .441, which, with chronological age constant, becomes .250. Since the sigma CAVI scores correlate negatively with age on account of the tendency for the younger children in any school group to be the brighter, the partial r between age and persistence, with the sigma score constant, rises from .331 (population Y) to .368. This figure shows the effect of increasing age on the tendency exhibited by children of the same brightness to persist at an intellectual problem. It should be compared with the r of .285, which shows the effect of brightness on the tendency exhibited by children of the same age to persist at an intellectual problem. On the other hand, the r of .202 shows the effect of age among children whose mental ability or maturity is the same, and the r of .250 shows the effect of mental ability or maturity among children whose chronological age is the same. It is clear that increasing age, greater native brightness, and maturing intellectual ability all tend to increase persistence scores on the Cross and Ring puzzles in population Y about equally.

changes with age and intelligence. They are consistently upward in X and Y, but on the whole downward in Z, which suggests that the associations with age and intelligence are probably due as much to experience as to inner growth.

Persistence in our tests is only a slight indication of what pupils will receive in scholastic marks or deportment. The average r's are .10 and .19 respectively.

The facts for inhibition are quite different from those reported for persistence. No consistent effect of retardation or acceleration is found, and age is almost wholly unrelated to score attained on the inhibition tests. The average r is .10. The correlations between intelligence and inhibition are far lower than in the case of persistence, reaching only .071 in X, .003 in Y, and .269 in Z. Evidently in X and Y the less intelligent are as well inhibited as the more intelligent.

Slight and consistent increases in inhibition appear from grade to grade, which are greater than might be expected from the correlations with age, suggesting that school experience probably has some effect on this type of behavior.

In population Z, teachers' marks for scholarship reflect the pupils' varying power of inhibition, as may be seen by the r of .431 between them and the Puzzles, Safes, and Stories tests combined. The average r for the three populations between marks and total inhibition scores is .15. In all populations good deportment tends to go with inhibition to the extent indicated by r's of .258, .314, and .306 for X, Y, Z, respectively, which average .29.

CHAPTER XXIV

SEX AND PHYSICAL AND EMOTIONAL CONDITION IN RELATION TO SELF-CONTROL

SEX

Sex and Persistence. Girls work harder than boys on the Story Resistance test and the Magic Square test, but there are no consistent differences in the others. The complete data are given in Tables LXXVIII to LXXXI of Appendix C. On the two tests most like school work, *viz.*, the Stories and the Magic Square puzzle, the difference between boys and girls is least in population X, which tends to bear out our suggestion that the boys and girls are more equally motivated where the schools use freer methods. Figure 82 shows the sex differences for the total persistence score. The girls of X are on the whole less persistent than the boys, but in Y and Z they are more persistent. The chances are a hundred to one that they would always prove so in population Z if the tests were repeated.*

In reputation for persistence, the girls enormously exceed the boys,† even in population X, where the boys had a higher score than the girls. As in the case of reputation for service, we find an apparent prejudice in favor of the girls, although in this instance the reputation score for persistence contains no pupil opinions at all. Since the tests and opinions deal with somewhat different modes of conduct, the sex differences in reputation may, of course, represent actual facts. Figure 82 is so drawn as to show the relative size of reputation in comparison with score. Not only do the shaded bars (girls) exceed the black bars (boys) in the lower part

* For all populations combined, girls exceed boys by 1.7 σ. In 96 out of 100 similar experiments the girls could be expected to be superior to the boys.

† The difference for all populations combined is 7.6 times its SE.

of the figure, which shows the facts for reputation, more than in the upper part of the figure, which shows the facts for scores, but the boys appear to have a lower reputation score than the tests would warrant, as may be seen from the fact that the black bars

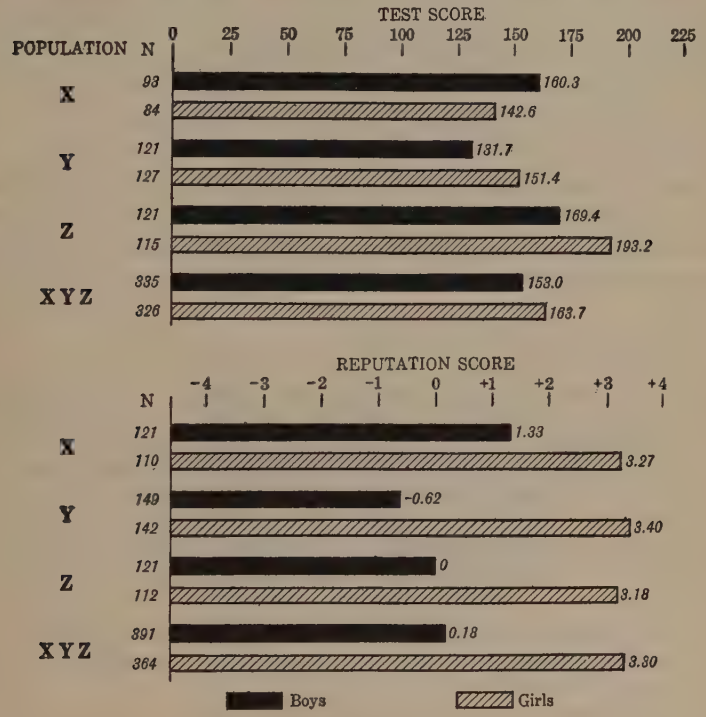


FIGURE 82

SEX DIFFERENCES IN MEAN PERSISTENCE SCORE (UPPER) AND MEAN REPUTATION SCORE (LOWER) — POPULATIONS X, Y, AND Z

of the lower figure are shorter than those of the upper figure. But in the case of the girls, the opposite is true — the shaded bars of the lower figure are longer than those of the upper figure.*

* The comparisons are made possible by the following device: First, the amount of difference between the sexes in reputation scores (about 3) is plotted

Before attempting any explanation of the sex differences in persistence we should have before us the facts for inhibition, to which we may now turn.

Sex and Inhibition. There is no doubt at all but that the girls are genuinely better inhibited than the boys in the types of conduct included in our tests. In all tests in each population, the differences favor the girls.* Full details are given in Tables LXXXII to LXXXVI of Appendix C. The total inhibition scores and the reputation scores are pictured in Figure 83.†

Not only are the total inhibition scores of the girls larger, but on the three tests where there is a clear demarcation between those who inhibit and those who do not, the per cent of girls who succeed in resisting the distractions presented is always greater than that of the boys, except for the Puzzle Manipulation test in population Y.

---

so as to occupy 4 times the linear distance representing the difference between the sexes in test scores (about 9), since the significance of the difference in reputation is about 4 times what it is in test score. Next, the position of the boys in reputation for X (1.33) was made to fall at the point on the scale corresponding to the average of the boys and girls together on test score (about 150). This point was chosen since the boys were apparently rated too low by about as much as the girls were rated too high. From this fixed point on the scale, with the units of about the size indicated, the remaining bars were plotted. The upper and lower figures may therefore be directly compared.

* The differences divided by their SE's for all populations combined are: Pictures, 2.36; Puzzles, 5.5; Safes, 4.8; Stories, 6.4; total inhibition score, 5.5; reputation for inhibition, 5.0.

† The reputation score is scaled to show the significance of reputation in comparison with test score. The difference between the sexes in reputation (3.3) is made equal to the difference in test score (2), since the significance of the difference in test scores just about equals that between reputation scores. Since the girls exceed the boys in test as well as reputation and since the difference between their test score is about 3 times as significant as in reputation, we may assume that there is a slight prejudice in favor of the boys on the part of the raters. Hence the reputation score of boys (2.40) is placed at a point slightly higher than the point corresponding to the mean of the boys in test score (21.6). With this unit and fixed point, the other bars were drawn. Hence the upper and lower figures may be directly compared.

As far as reputation goes, the differences are about as significant as for the total score,* indicating less prejudice with respect to reputation for this type of conduct (or greater knowledge) than

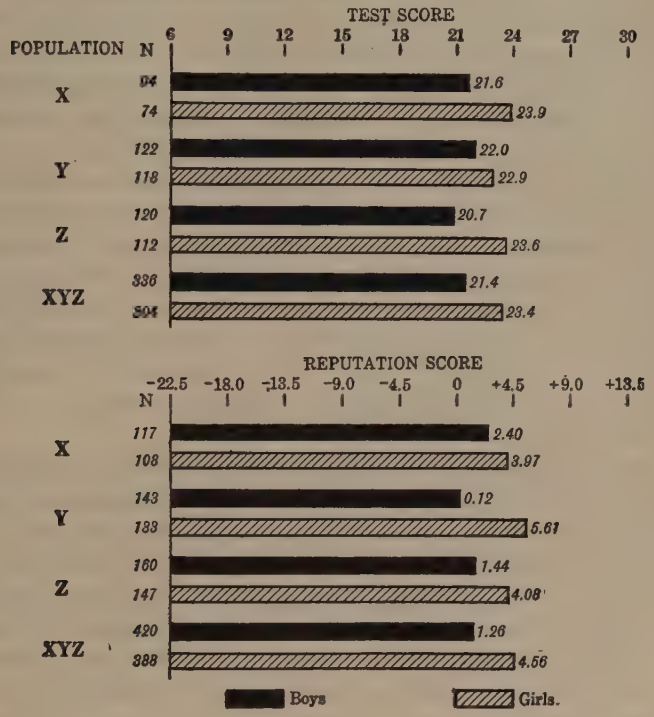


FIGURE 83

SEX DIFFERENCES IN MEAN INHIBITION SCORE (UPPER) AND MEAN REPUTATION SCORE (LOWER) — POPULATIONS X, Y, AND Z

in the case of persistence and coöperation. This may be seen by comparing the upper and lower graphs of Figure 83. The differences between the black and shaded bars are about the same for the test and reputation scores.

---

* See footnote on preceding page.

POSSIBLE EXPLANATION OF SEX DIFFERENCES IN SELF-CONTROL

As in our discussion of the older in comparison with the younger children, we find the superiority of girls over boys in persistence to be associated with a corresponding superiority in inhibition. Here, however, the bigger differences are in the inhibitive behavior rather than in persistence, whereas the reverse is true of the older as compared with the younger children. Consequently we cannot attribute the superior persistence of the girls to greater interest on account of the fact that such hypothetically greater interest would tend to make inhibition more difficult for them, whereas they actually are better inhibited. The girls, to be sure, are but little more persistent than the boys, so that, even if the difference be due to interest, there is still room for the observed difference in inhibition. Indeed, any differences in interest that might be demonstrated only serve to emphasize the superiority of the girls in power of inhibition.

If we compare the relative interests of the two sexes in the " temptations " of the inhibition tests, we would naturally expect to find boys showing greater avidity for puzzles and safes. This assumption is supported by the fact that on the Cross and Ring test the boys show more persistent conduct than the girls, and on the Puzzle and Safe Manipulation tests the difference between the two sexes in power to resist the desire to handle these attractive objects is more than twice as great as in the case of the Picture distraction test.

The superior inhibition of the girls on the Picture test may also be a function of their possible greater desire for good marks. As we noted in our *Studies in Deceit*, girls tend to cheat more than boys on tests taken home, which we attributed to a difference in " school drive." The same factor would tend to appear in the inhibition of any distractions which might interfere with their making a good record. It ought also to operate to make them especially superior on tests most similar to school demands, such as the Persistence for Self test, but as a matter of fact there is little difference on this

test between boys and girls. We can account for this, though, by the fact that there is here an element of monotony and fatigue, which may give the boys enough advantage to enable them to overcome their handicap in lack of interest.

We have referred to the tendency for boys of more progressive schools (perhaps because of selection) to be more nearly equal to the girls in school drive than in the general run of schools, and we have also had occasion to point out that speed of work is partly a function of inhibition of distractions. It is interesting in this connection to note that the difference between boys and girls in power of inhibition is far less in the X population than in the other two, that in the self and class persistence scores the X boys are actually superior to the X girls, and that in general level of speed ability they more nearly approach the girls than in either of the two other populations.

PHYSICAL CONDITION

Since our tests of physical condition are largely measures of effort to overcome fatigue, one might think that there should be some association between scores on these tests and on our tests of persistence and inhibition, but the fact is that our persistence total shows absolutely no relation to physical condition as measured by our tests. Our Persistence for Class and Persistence for Self tests, however, also involve the factor of fatigue, and here we might expect something in the way of association, but the *r*'s in populations X, Y, and Z are all approximately zero.

The case for inhibition is much like that for persistence. Here the *r*'s run .173, .132, and .076 for populations X, Y, and Z. If we consider only X and Y, we find that the children who score extremely high for their age on our physical ability tests are also well inhibited. Of the 17 cases scoring 120 or over (100 is the average) on the physical condition battery, only two are below the median for their school.

The Health Examination. We have also for population Y the records of the health examination made by the American Child

Health Association, which we described in Chapter VIII. The results of this examination, when expressed as a total score, correlate .11 with persistence. Figure 84 shows the facts for vision, nutrition, and condition of teeth. It is interesting that eight children having extremely poor eyes scored 210 as against a median of 126 for the whole group, which is a difference of about $\frac{1}{5}$ of the whole scale. Those of average eyesight had average persistence scores.

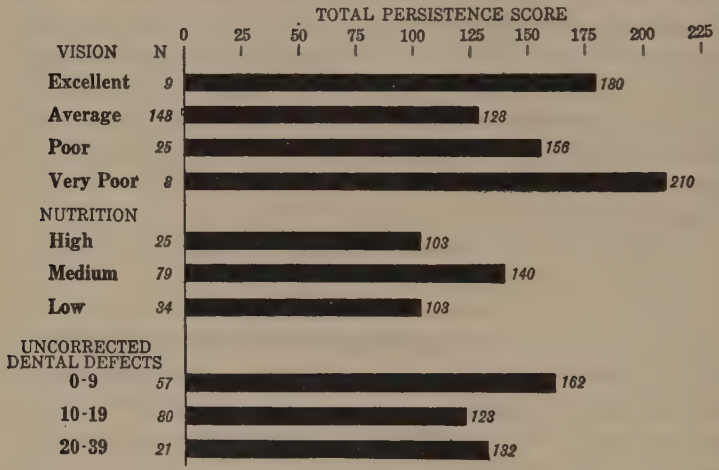


FIGURE 84

VISION, NUTRITION, AND CONDITION OF TEETH IN RELATION TO PERSISTENCE

Also those in the middle of the nutrition scale had the highest average persistence score, with the best and worst nourished falling far below. In the case of teeth, the third factor, evidence of uncared-for defects and evidence of well-tended teeth both were associated with relatively high persistence, whereas those with average teeth had lower scores.

As to inhibition, its correlation with the sum of the three health tests is zero, and its relations with vision, nutrition, and condition of teeth are all negligible.

EMOTIONAL CONDITION

In measuring emotional stability we used the Woodworth-Mathews Personal Data Sheet, which is described on page 163. The score represents the number of questions which are answered in such a way as to be symptomatic of neurotic condition. The lower the score, the greater the emotional stability.

Offhand one might expect that persistence and inhibition would be associated with the poise and balance of personality represented

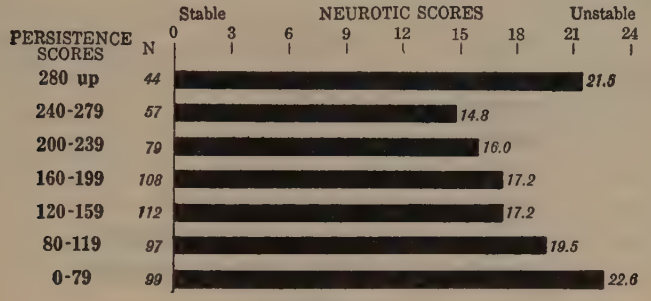


FIGURE 85

MEDIAN NEUROTIC SCORES FOR VARIOUS LEVELS OF PERSISTENCE — POPULATION XYZ

in low scores on this test. We find this to be rather generally true in the case of persistence, but less true in the case of inhibition.

Persistence. Expressed in terms of correlation, the coefficients are .23, .15, and .08 for populations X, Y, and Z, but these are not as valuable * indices of the true association as are the medians for the several levels of persistence, which are summarized in Figure 85 for all populations combined.† Save for the very persistent pupils, there is a steady decrease in stability with decrease in persistence.

So far as nail biting is evidence of nervousness, the facts regarding the practice correspond to those just reported. Those who do

* The regressions are curvilinear.

† That is, the medians are unweighted by their respective N's, as each population is taken as a unit.

not bite their nails have a mean persistence score of 156.5, and those who do bite their nails have a mean score of 116.5.* This is a substantial difference in favor of those who do not have this habit.

Inhibition. The correlations between emotional stability as measured by the Woodworth-Mathews device and inhibition are — .066, .08, and .305 for populations X, Y, and Z. The average is .11. In the Z population alone is the relationship significant. In population Y, nail biters score 11.5, and those who do not bite their nails, 13.2.

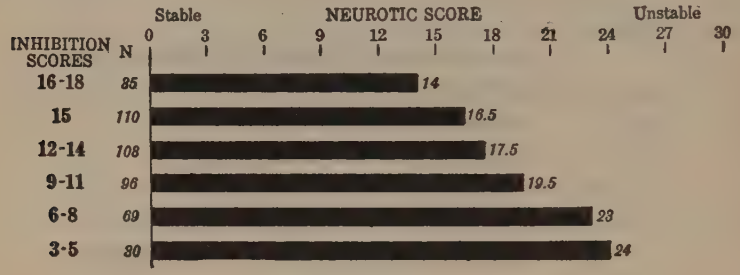


FIGURE 86

MEDIAN NEUROTIC SCORES FOR VARIOUS LEVELS OF SCORE ON THREE INHIBITION TESTS — POPULATION YZ

By confining attention to the Safes, Puzzles, and Story Inhibition tests we come upon some interesting relationships. Since the range of symptoms in X is very narrow and since it appears that inhibition scores are unaffected unless the symptoms average greater than around 12, we may leave population X out of account and consider only Y and Z. When the three tests are combined, in case an individual passes each test perfectly or nearly perfectly, his score is 16, 17, or 18. These three scores are therefore grouped. In case he resists two of the three test stimuli, his score is 15; therefore 15 is retained as a separate step. Lower scores are grouped by threes. Figure 86 shows that those of Y and Z who

* The difference in mean persistence scores is 4.7 times its SE. The biserial *r* between persistence and nail biting is — .28.

resisted all three test stimuli perfectly or nearly so have on the average around 14 symptoms of instability, which, according to the Mathews' norm, is 9 points below normal. But, on the other hand, those that failed on all three tests have on the average about 24 symptoms, which is one point above normal. In between these extremes the average number of symptoms increases with increasing failure to resist the test stimuli.

The Woodworth-Mathews test has to do mainly with social adjustments and maladjustments. Our tests require an impersonal sort of self-control. On the manipulation test, for example, the child is expected to inhibit the tendency to touch an object. This is quite different from inhibiting the tendency to strike when angry or to speak when provoked. If these tests involved the manipulation of persons rather than toy safes or puzzles, the correlations with the Woodworth-Mathews test would probably be much higher.

CHAPTER XXV

GENERAL ENVIRONMENT AS A FACTOR IN SELF-CONTROL

COMMUNITY DIFFERENCES

In Chapter IX we described the three communities to which our present school populations belong and pointed out several respects in which they differ from one another. Population X is in a residential section of New Haven; population Z is in a district in the same city the members of which are largely foreign born; population Y is in Walden, New York, which is a town of some 5000 population, containing its own industries and entirely surrounded by farms. Table XXX summarizes the chief points of difference and shows that population X is highest, Y next, and Z lowest in such matters as intelligence, emotional stability, social and economic status, school marks, and deportment; but with Z somewhat older on the average and also of greater athletic ability, age for age. Such community differences as are represented by the Burdick culture scores and Sims socio-economic scores, described in *Studies in Deceit*, and by occupational level are shown in Figures 33, 34, and 35.

In Book One we found that the children of Y were the most helpful of our three populations, with X standing second and Z third. This was somewhat of a surprise, as one might think that X should have scored highest on our service tests. It may be still more of a surprise to find that in persistence X again fails to reach the top and is only slightly better in inhibition than Y and Z. For persistence the order of populations is Z, X, Y, and for inhibition X, Y, Z. The details are of sufficient importance to demand our attention for a moment.

Population Differences in Persistence. The percentage distribution of total scores is shown in Figure 87. It may be seen that the solid line, representing X, is between the broken line (Y) and the dotted line (Z), with the latter toward the upper end of the scale. The details are given in Tables LXXXVII to XCII of Appendix C. The difference between Z and Y is nearly three times

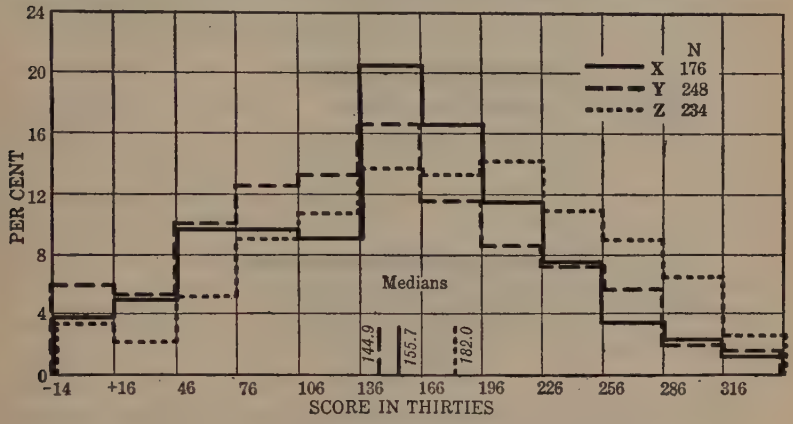


FIGURE 87

PERCENTAGE DISTRIBUTION OF TOTAL PERSISTENCE SCORES BY POPULATIONS

as great (in statistical significance) * as the difference between X and Y.

We have already considered the possibility that Z may rank highest in persistence because of a tendency to dawdle on the Ring and Cross and Magic Square tests, on which they show bigger scores than the rest, attributing this to their relatively lower intelligence and lack of interest. But we were able to show that these tests correlate positively, not negatively, with intelligence, so that the expectation should be the reverse of what we find. These and other test differences are all pictured in Figures 88 and 89, with

* X − Y is 1.35 times its SD. Z − X is 3.78 times its SD. Z − Y is 5.45 times its SD.

X, as usual, shown by a solid bar, Y by a crosshatched bar, and Z by a shaded bar. The results may be summarized by the follow-

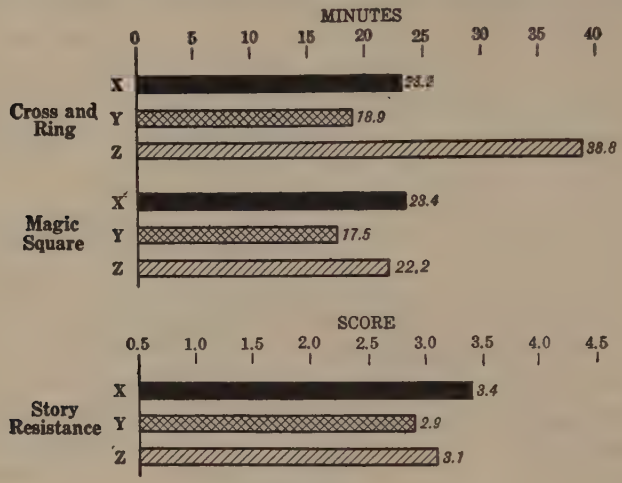


FIGURE 88

Population Differences in Median Scores on the Cross and Ring, Magic Square, and Story Resistance Tests

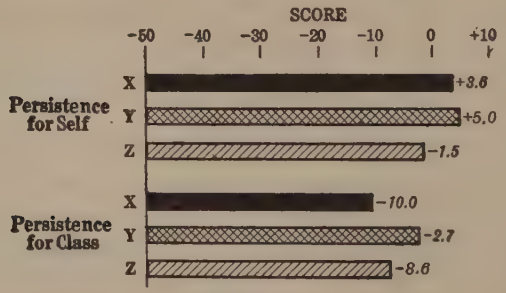


FIGURE 89

Population Differences in Median Scores on the Persistence for Self and for Class Tests

ing table, which assigns first, second, or third place to each population on each test:

Test	Position		
	Population X	Population Y	Population Z
Story Resistance	1	3	2
Cross and Ring	2	3	1
Magic Square	2	3	1
Persistence for Self	2	1	3
Persistence for Class	2	1	3
Total firsts	1	2	2
Total seconds	4	0	1
Total thirds	0	3	2

Population Differences in Inhibition. The percentage distribution of total scores would look very much like Figure 87. Again the overlap is very great. X ranks only slightly higher than Y, and Y only slightly higher than Z.* Figure 90 pictures the score of each population on each of the four inhibition tests. The details are given in Tables XCIII to XCVIII of Appendix C. We may summarize these tables as follows:

Test	Position		
	Population X	Population Y	Population Z
Picture Inhibition	3	2	1
Puzzle Manipulation	2	1	3
Safe Manipulation	2	1	3
Story Inhibition	1	3	2
Total firsts	1	2	1
Total seconds	2	1	1
Total thirds	1	1	2

Evidently the small differences in total score are due to the specific ways in which the populations respond to the situations of the separate tests. When it came to deciding whether to finish a ghost story or count letters on a page in order to make a score, it

* $X - Y = 1.44\,\sigma$. $X - Z = 1.01\,\sigma$. $Y - Z = .69\,\sigma$.

was the pupils of population Y who yielded by far the most frequently to the temptation to read instead of count.* On the Picture test, when it was a choice between working on the additions or looking at pictures and other interesting things, the pupils of both X and Y proved far less able to resist temptation

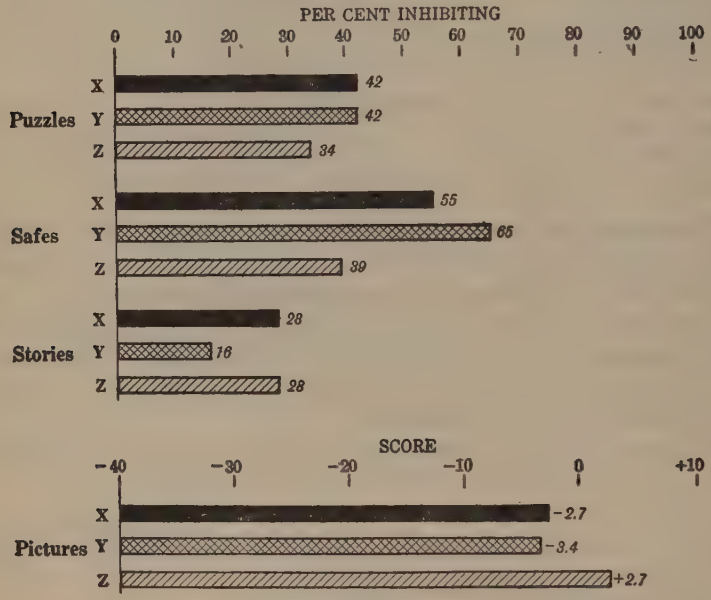


FIGURE 90

POPULATION DIFFERENCES IN CONDUCT ON THE FOUR INHIBITION TESTS

than those of Z.† In ability to let the safes alone, it was the pupils of Z who were least controlled, with X next and Y at the top.‡

* The difference in percentages succeeding in inhibiting the desire to read is 28.2 − 15.6, or 12.6, in favor of X, and 28.0 − 15.6, or 12.4, in favor of Z. Each of these differences is about 3.5 times its SE.

† The mean scores of X and Y are practically identical. Z − X is 4.9 σ. This difference is genuine and not due to the nature of the Picture Inhibition score. See Appendix G, 2.

‡ The per cents succeeding are: for X, 55; for Y, 65; for Z, 39. The differences in terms of their SD's are: Y − X, 2.3 σ; Y − Z, 6.6 σ; X − Z, 3.7 σ.

Finally, on the Puzzles test, the pupils of Z, again, were far less able to resist the puzzles than the children of Y and X, and Y was again at the top.* Thus we have extreme differences between populations on separate tests, with first one population ahead and then another. This crisscrossing of differences evens up the total scores.

Summary of Community Differences in Self-Control

In persistence, population Z, the least favored of the three communities, is so superior on two out of five tests as to raise its total persistence score far above that of either X or Y. These two tests are the Ring and Cross puzzles and the Magic Square puzzle. Either these children do not realize when they are beaten or they have greater interest in this kind of problem. Certainly we cannot attribute to them any *general* tendency toward superior persistence, as they fall below (though not much below) both X and Y on the Persistence for Self and Persistence for Class tests, which require application to a fatiguing task, and below X on the Story Resistance test, which also soon gets wearisome.

Similarly, we find no general population differences in inhibition, but extreme differences on separate tests, each population proving superior on at least one test and inferior on at least one, thus evening up the total scores. X's superiority is on the Story Inhibition test, on which it breaks about even with Z. Y stands out on the Safes and Puzzles. Z is ahead on the Pictures, sharing the honors with X on the Stories, as just noted.

OCCUPATIONAL DIFFERENCES

Following the procedure of Book One, we shall deal separately with such aspects of the general environment as we have measured or for which we have any indicators. The first and most accessible of these is the father's occupation. For purposes of comparison we used, as before, the classification adopted by Sims,† level I being

* The difference between Y and Z is 5.8 σ in favor of Y.

† The Sims Score Card, Public School Publishing Company, Bloomington, Ill.

the highest. In contrast with our findings for service tendencies, no consistent relation exists between occupational level and persistence. Beginning with the top level, the median persistence scores run as follows: for X, 150, 137, 154, 170, 170; for Y, 230, 130, 147, 144, 147; for Z, beginning with level II, 233, 152, 190, 185.

Much the same is true of occupation and inhibition, though the upper levels of X and Z score distinctly above the lower. Beginning with the top level, the median inhibition scores run as follows: for X, 22.3, 23.3, 24.0, 20.5; for Y, 23.5, 22.0, 22.9, 22.2, 24.3; for Z (beginning at II), 24.2, 24.5, 21.6, 21.5.

ECONOMIC AND CULTURAL DIFFERENCES

The Burdick culture test, which attempts to measure general cultural level by securing the reactions of the pupils to a great variety of imaginary situations, was given to all three populations. The status of the three groups is shown in Figure 33, page 173.

Culture and Persistence. The correlations between persistence and the scores on the culture test are .146 for X, .248 for Y, and .148 for Z. These, though small, are positive and indicate a slight trend within each population for culture and persistence to go together. It is sufficient, however, to demonstrate that it is not because of the presence or absence of cultural advantages that population Z is superior in persistence and Y inferior, important as these factors may be within any one community.

Culture and Inhibition. The total inhibition score shows a small but positive relationship with culture score, with an r of .097 for X, .132 for Y, and .322 for Z.*

Socio-economic Status. Similar computations were made for the Sims score of socio-economic status. The r's are $-$ .022 for X and $-$ .011 for Y, in the case of persistence, and .026 for X and $-$ .100 for Y in the case of inhibition.

* It is interesting to note that the correlation between the Woodworth-Mathews neurotic index and the Burdick score is $-$ .32 for X, $-$ .28 for Y, and $-$ .32 for Z, which is consistent with the facts for inhibition.

Summary of Occupational and Cultural Differences in Self-Control Scores

Such economic and social facts as are represented in occupational level, the scores on the culture test, and the answers to straight questions about living conditions are only slightly related to persistence and inhibition. The Burdick test reveals the most consistent relations, but even here the correlations show that there is only the loosest connection between the types of conduct we are dealing with here and the social status of the subjects. Within any one community, the children of poorer parents, whose families have fewer cultural opportunities and whose fathers occupy humble or subordinate positions, are on the whole only slightly less inclined to stop a task before it is finished and only slightly more apt to be distracted from duty by irrelevant temptations than are the children of the well-to-do and the cultured who hold positions of skill or responsibility. Certainly, then, the differences between X, Y, and Z in persistence and inhibition cannot be accounted for by these associations between cultural status and these behavior tendencies.

NATIONAL DIFFERENCES

Our major populations vary considerably in the proportion of children whose parents are foreign born. Consequently, it is essential to find out, if possible, whether there is any association between nationality and such conduct tendencies as are recorded in our test results. It may well be that the differences we have noted between X, Y, and Z are due largely to the influence of national heterogeneity. Population Z, for example, is both more persistent and more foreign than the others. If children of foreign-born parents in all populations show superiority on this test, this would account in part for Z's position in relation to X and Y; and this, in fact, turns out to be partially true. The children of foreign-born parents, on the whole, have bigger persistence scores than the children of native-born parents and bigger inhibition scores on the Picture Inhibition test. This is the

first type of behavior on which this group has been so far found to excel, and we must therefore look into the matter more closely.

Nationality and Persistence. Let us turn first to the total persistence score, comparing Figure 44 with Figure 91. In each population except Z, the children of the foreign born show higher persistence scores than those of the native born. Indeed 73% of those

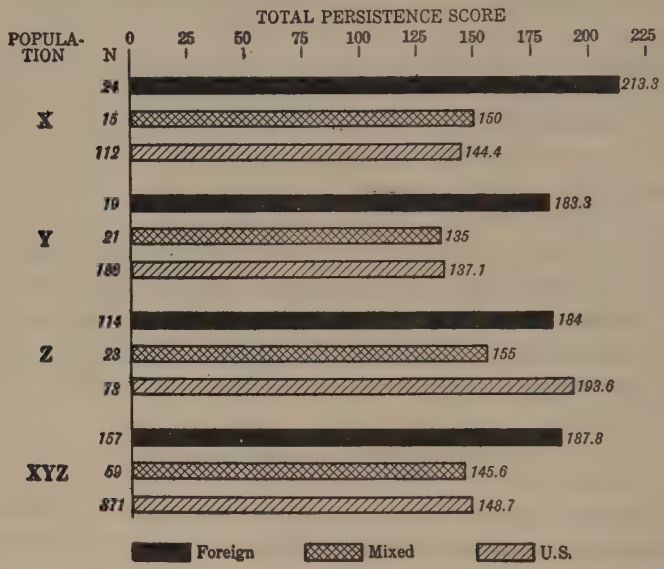


FIGURE 91

COMPARATIVE PERSISTENCE SCORES OF CHILDREN OF NATIVE- AND FOREIGN-BORN PARENTS—POPULATIONS X, Y, AND Z

having foreign-born parents equal or exceed the average persistence score of children of native-born parents.* The Italian group of all three populations, furthermore, is superior to the native group, and the Irish is superior in X and a close runner-up in Y and Z. Thus the facts for service are almost completely reversed. But it should also be noted that the children of native-born parents in

* See Table XCIX, Appendix C.

Z are *also* superior to the native group of X and Y, as well as to those of their own schoolmates whose parents are foreign born. So, even though the foreign group tends on the whole to excel the native, which alone would account for Z's superiority, we must look also for other factors since the native group of Z is even more persistent than the foreign.

Nationality and the Separate Persistence Tests. Further light is thrown on the influence of nationality by an examination of the separate tests. These we have treated in three groups: one consisting of the Ring and Cross test and Magic Square test combined, on which Z showed its extreme superiority; one the sum of the Persistence for Self and for Class tests, involving monotonous work, and the Story Resistance test, which is a sort of puzzle. The facts are presented in Table C, Appendix C. The superiority of the foreign groups on the first combination is clear enough: 69% of the foreign group equals or exceeds the median of the native group.* Furthermore, the Z groups of foreign, mixed, and native parentage are in each case superior to the X and Y groups of like parentage.

The Story Resistance test reveals similar superiority on the part of the foreign groups of each population. The amount is slight but consistent. And the Italians and Irish again excel. But in the case of the tests involving monotonous additions the foreign and mixed groups are on the whole slightly less persistent than the native, although the Irish have the highest scores of all.

Thus it turns out that population Z, which is largely foreign, is equal or superior to Y and X on those tests on which the foreign group of *each* population most conspicuously exceeds the native.

In general reputation the native group exceeds the foreign in both persistence and inhibition, as shown in Figure 92. That this opinion is not entirely due to prejudice against the children of foreign-born parents is seen from the fact that the Italian group slightly exceeds the general foreign average in reputation for both persistence and inhibition and from the fact that the Irish group

* The difference between the means is 4.6 times its SE.

has a reputation score for inhibition far above that of even the native group.

Nationality and Inhibition. Turning now to the consideration of inhibition scores, we find the facts reported in Figure 93. In populations X and Y, the children of the foreign born have higher scores — that is, are better inhibited — than the children of the native born. This corresponds to what we found for persistence. Furthermore, the Italian and Irish groups of Z equal or exceed the

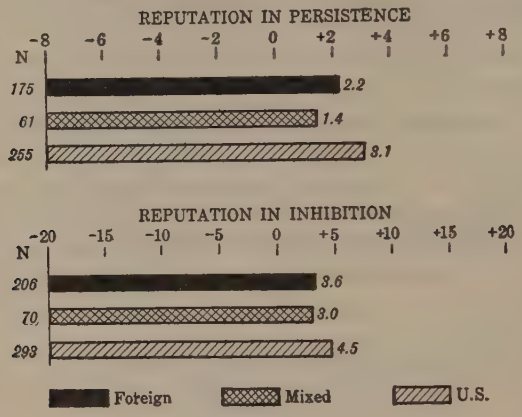


FIGURE 92

COMPARATIVE SCORES IN REPUTATION FOR SELF-CONTROL OF CHILDREN OF NATIVE- AND FOREIGN-BORN PARENTS — POPULATION XYZ

average of all the foreign, just as in the case of persistence. When all the foreign of all populations are compared with all the native, the two groups are found to have the same median inhibition score.

Considering the Picture test alone, the children of parents born in other countries again are better able to resist the distracting stimuli than those whose parents were born in the United States. This is true even in population Z, where the native group is usually superior to the foreign. Combining the populations, 63% of the foreign group equal or exceed the median score of the native group. On the other three inhibition tests combined, the native group is,

on the whole, slightly superior. These facts and all relevant data are given in Tables CI and CII of Appendix C.

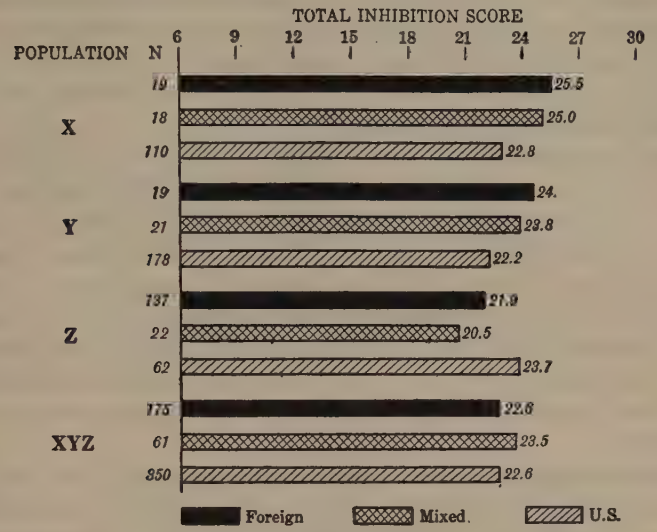


FIGURE 93

COMPARATIVE INHIBITION SCORES OF CHILDREN OF NATIVE- AND FOREIGN-BORN PARENTS — POPULATIONS X, Y, AND Z

RELIGIOUS DIFFERENCES

It is commonly believed that a child's conduct is somehow determined, in part at least, by the religious background or the religious practices of parents. A child brought up in a strict Puritanic atmosphere might be expected to be drilled more in self-control and self-denial than one brought up by a parent whose philosophy and religion is more Epicurean. Or, again, the religious practices of Jews might very well have different effects from those of Catholics. Just what these differences are, no one knows. Neither does anyone know the extent to which any kind of conduct is affected by any kind of religious belief, practice, or general atmosphere. With

the advent of objective character tests we shall soon be in a position to find out the nature and extent of whatever effects there may be. Are Jewish children more persistent, more long-suffering, in accordance with the popular conception of their racial characteristics? Or do Protestant children have a more variable ethical outlook than Catholic children? To such questions as these science will soon be able to give a definite answer.

The findings of the present studies are, of course, too meager to be regarded as anything final. They are reported more as suggestive of what can be done provided these or similar tests are given to representative groups. The indication of our sample study is that the Jewish children average higher scores in both the inhibition and the persistence tests than either the Protestants or the Catholics. But the Jewish group which is compared with the others contains only a few cases, and even the larger number of Protestants and Catholics is not enough to afford a basis for generalization.

Religion and Persistence. The differences among the religious groups in the tendencies to persist which we have measured are shown in Figure 94. The Jews have the highest total persistence scores, and the Catholics come next in X but have low scores in Y. Combining all populations, the order is Jews, Catholics, Protestants.*

These facts refer to the total score. In population X, where we had a fair representation of each of the three religious groups, we compared the scores for the different types of behavior. It turns out that the Jews are superior † only on the puzzle tests and are about equal to the Protestants on the Story Resistance test and the two tests which involve monotonous additions. On the latter the Catholics excel,‡ as also on the Story test.

---

* In population X, the Jews exceed the Catholics by 1.1 times the SE of the difference and the Protestants by 1.5 σ. The Catholics exceed the Protestants by .47 σ. When all are combined, the Jews and Catholics exceed the Protestants by .78 σ and 1.7 σ, respectively.

† .75 σ over Protestants and .6 σ over Catholics.

‡ .59 σ over Protestants and .81 σ over Jews.

Religion and Inhibition. Before drawing any conclusions, it would be well to have in mind also the results of the inhibition tests. The medians on the total score for each group and each population are shown in Figure 95. Here again, the Jews are slightly superior,* and, except for population X, the Catholics come next.

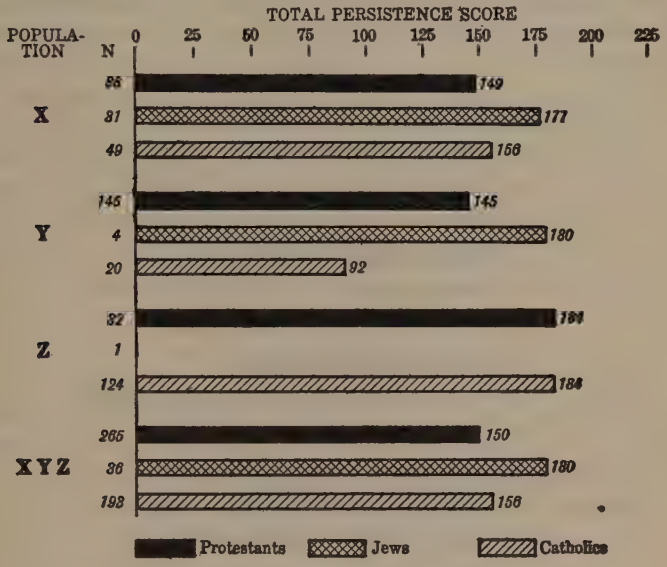


FIGURE 94

COMPARISON OF MEDIAN PERSISTENCE SCORES OF CHILDREN WHOSE PARENTS ARE AFFILIATED WITH THE THREE RELIGIOUS GROUPS — POPULATIONS X, Y, AND Z

These differences, however, are not substantial enough to afford a basis for generalization even about these populations, much less about the three religious groups at large. Taking the Picture test by itself, the Catholics do best in Y and Z and almost equal the

* The differences between the means are all small. Combining the three populations, the difference between the Jewish and Catholic groups is 1.4 σ and between the Jewish and Protestant groups, 1.2 σ. In Z alone, the Catholics exceed the Protestants by 1.5 times the SE of the difference.

Protestants in X, where the Jews show the highest scores. The sum of the other three inhibition tests shows only slight differences among the three religious groups.

Religious Differences in Reputation. One might expect prejudice to influence judgment when religion is involved; and since most of the teachers concerned were Protestant, any such prejudice should operate to place the Protestants higher and the other

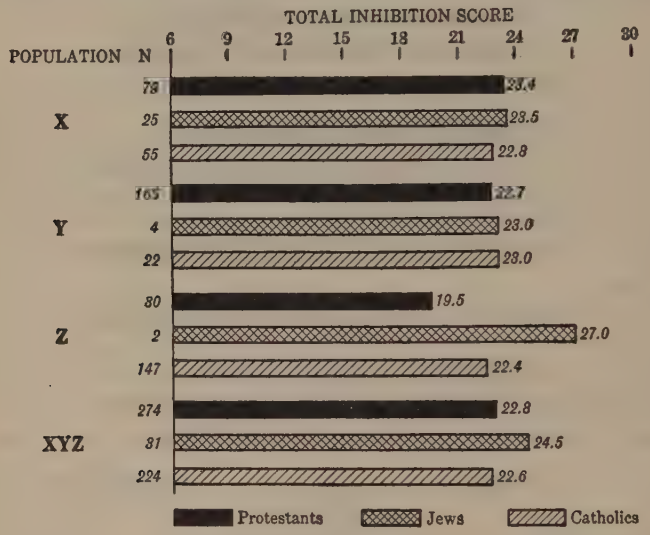


FIGURE 95

COMPARISON OF MEDIAN INHIBITION SCORES OF CHILDREN WHOSE PARENTS ARE AFFILIATED WITH THE THREE RELIGIOUS GROUPS — POPULATIONS X, Y, AND Z

two groups lower than their actual behavior might indicate. This cannot be exactly checked by our scores, since the behaviors required by the tests were by no means exactly comparable to the conduct included in the reputation scores. It is of interest, however, that in population X the reputation of the pupils among the teachers agrees with the scores in placing the Jews first in persistence; but the Catholics are rated far below the Protestants,

whereas they scored higher than the Protestants.* In the matter of inhibition, the Protestants are rated the highest, the Jews next, and the Catholics, again, far lower.* The scores, it will be recalled, are closely alike for the three groups.

Summary of National and Religious Differences in Self-Control

National background evidently exerts considerable influence in the matter of persistence and inhibition, just as we found to be the case in service tendencies. In the forms of behavior we have called self-control, the differences frequently favor the children of the foreign born. No corresponding differences in religious background are observable, however, nor is the frequent superiority of the foreign group attributable to cultural differences, for these on the whole would favor the children of the native born. Differences in intelligence, furthermore, would lead one to expect opposite results from those found. One is tempted to attribute the superior control of the foreign group to some generalized compensation for repression in a new country or to a hypothetical generalized result of discipline through hardship or to some general result from release of energies in a land of opportunity, but it is more likely that the differences reflect the operation of subtle factors in the control of each form of conduct.

CONCLUSIONS CONCERNING GENERAL ENVIRONMENT AS A FACTOR IN SELF-CONTROL

Enough has been said in the preceding summary to suggest the trend of our conclusions regarding the influence of environment. Our populations differ from one another on the various forms of behavior represented in our several tests of self-control. These

* In reputation for persistence the Protestants exceed the Catholics by 4.3 times the SE of the difference, whereas the difference in test scores is a half sigma in favor of the Catholics. In reputation for inhibition, the Protestants exceed the Catholics by 3.4 times the SE of the difference, whereas their mean test score is only slightly higher.

differences are often extreme, but they are not general. They correspond roughly to the differences noted in national constituency, the population with the largest proportion of foreign stock proving superior on the tests on which all children of foreign-born parents tend to excel. Hence we cannot attribute the differences solely to the peculiar influence of the different communities. Nor does social-cultural level produce any general effect comparable to these other group differences, for, although the children of the well-to-do are usually better controlled than those of lower socio-economic levels in their own neighborhood, yet the foreign group, which has low economic status, has the higher control scores on many tests.

Religious affiliation plays only a small part in the picture; but, so far as it enters at all, it serves only to complicate the causes determining tendencies to work long and hard at assigned tasks and to resist temptations to be distracted from them. The net result is a series of trends specifically related to combinations of subtle factors in the situations rather than to any general standard or ideal or social demand.

CHAPTER XXVI

HOME AND SCHOOL IN RELATION TO SELF-CONTROL

We have not made a direct comparison of the amount of influence exerted by the school with the amount exerted by the home. School and home are constantly at work on the child, so that it is almost impossible to disentangle the influences which have their origin in these two institutions. We shall do the best we can, however, to present clear-cut material even though there may be differences of opinion regarding its interpretation.

THE HOME

Since the time when Galton established the fact that genius runs in families, many researches have shown how the same traits appear from generation to generation, tending to make children resemble their forebears and relatives more than they resemble the offspring of other stocks or the subjects of other cultural influences. In the case of traits of conduct, as compared with structural differences, it is easier to establish the fact of family resemblance than the certainty of biological determination. Indeed, the social and biological influences are so interwoven that many have given up the task of attempting to separate them. Certainly the mere fact of resemblance does not establish the fact of a common inheritance, as we have repeatedly pointed out.

We shall present two types of data, one having to do with the resemblance of brothers and sisters and the other with the association between the intelligence of parents and the conduct of the children.

Sibling Resemblance in Persistence. We have described fully elsewhere* the techniques used to show the extent to which brothers and sisters resemble one another in any characteristic. Briefly, we take the family as a unit, the score of one sib as one measure of the family and the score of the other sib as the second measure of the family. These two sets of scores are then correlated. If the r should be 1.00, it would mean that brothers and sisters were exactly alike. If it should be about zero, it would mean that they resembled one another no more than they resembled children of other families.

The results of these computations are given in Table LIX.

TABLE LIX

CORRELATIONS BETWEEN SIBLINGS IN THE PERSISTENCE TESTS

TEST	N	r	$\sqrt{r}$
1. Story Resistance	204	.287	.536
2. Cross and Ring	204	.223	.472
3. Magic Square	208	.136	.369
4. Persistence for Self	188	.173	.416
5. Persistence for Class	176	.242	.492
2 and 3	170	.458	.677
4 and 5	150	.346	.588
All combined	130	.406	.637
Reputation	150	.306	.553

Intelligence and total persistence (sibs only), .055
Culture and total persistence (sibs only), .129

Comparison with Table XXXII shows that these r's are about what we found for service tendencies and indicate unquestionably that persistence of the sort measured by our test does more or less run in families.

As we have previously stated, these r's are produced by the fact of family variation. They do not in themselves indicate the cause of this variation. So far as the r's themselves are concerned, they might as well be due to differences in culture as to differences in

* See page 201 and also *Studies in Deceit*, Book One, Chapter X.

stock.* In *Studies in Deceit* we showed that, in order to explain the observed resemblances on environmental grounds, it would be necessary for the correlation between the environment and the characteristic in question to equal the square root of the observed correlation between the siblings. The right-hand column of Table LIX gives these square roots. But at the bottom of the table is the actual correlation between the Burdick culture score and the total persistence score (.129), which is far from equaling the *r*'s in the last column of the table.

If common family experiences are wholly responsible for the tendency for brothers and sisters to be alike on the persistence tests, then these common experiences do not leave their records in such measures of their influence as we have been able to use.

The correlation between intelligence and persistence is practically zero; so the resemblance of siblings in persistence could not be accounted for by their resemblance in intelligence.

Sibling Resemblance in Inhibition. The data for inhibition are similar to what we have reported for persistence. Table LX gives the essential facts. The *r*'s for the tests run about the same, but the coefficient for reputation is higher than in the case of persistence.

TABLE LX

CORRELATIONS BETWEEN SIBLINGS IN THE INHIBITION TESTS

TEST	N	*r*	$\sqrt{r}$
1. Pictures	196	.323	.568
2. Puzzles	170	.205	.453
3. Safes	196	.344	.586
4. Story Inhibition	196	.189	.435
2, 3, and 4	136	.251	.501
Reputation	222	.443	.665

Intelligence and 2 + 3 + 4 (sibs only), .013
Culture and 2 + 3 + 4 (sibs only), .141

* If due to differences in stock, they would be higher if each instance of inherited antithesis could be treated in the same way as each instance of in-

Again, the extent to which environment, as measured by the Burdick test, fails to account for the resemblances observed is shown by the differences between the coefficients of the last column and the r of .141, which is the observed correlation between the culture test and the sum of three of the inhibition tests.* And again, community of intelligence of course does not account for the resemblance actually found.

We may conclude, as we did in the case of service tendencies, first, that the resemblance of siblings in self-control is not quite so high as in the case of deception or intelligence and, second, that it is not to be accounted for by such environmental influences as we have measured.

Parental Intelligence and Self-Control. The facts bearing on the relation between the intelligence of parents and the self-control of children are briefly reported. Expressed in the form of correlations, the r's for the total persistence score are .242, — .069, and .301 for populations X, Y, and Z and .066 for all combined. For the sum of three inhibition tests, the r's are .169 for X, .067 for Y, .407 for Z, and .177 for all combined. Apparently, it is in the community with the lowest intelligence and generally unfavorable surroundings that such family differences as are associated with the intelligence of the parents are of most consequence. But these r's are all somewhat unreliable on account of the small number of cases, so that too much stress cannot be placed on these differences in their size.†

THE SCHOOL

In our study of service tendencies we were confronted with the problem of explaining why population Y, which in most respects

---

herited resemblance. If the facts were as clear as eye color or hair color, this might be done in a contingency table.

* The other methods available for measuring environmental influences showed still lower associations with self-control than did the Burdick test.

† The number of entries for the r's quoted is, for persistence, 65 in X, 43 in Y, and 29 in Z; for inhibition, 47 in X, 46 in Y, and 35 in Z. That is, we had intelligence scores for at least one parent for about 65 cases in X, 43 in Y, etc.

fell considerably below X, should be superior to X in helpful and coöperative behavior. The facts we have been reporting about persistence raise a similar problem, but in this case it is Y which is inferior and Z, the lowest in most characteristics, which is superior. In the case of service we felt that in the school situation of Y there were certain features leading to homogeneity of attitude and providing the conditions under which children might be expected to become increasingly coöperative. These conditions were not present in X or Z.

The contrasts between populations in their conduct on various tests of persistence and inhibition are so great and also so inconsistent with one another that it is impossible for us to point to any one cause or group of causes which may be operating to make the pupils of one population do better than the others on some tests and worse on some others. Some of these contrasts, however, are intelligible, and we shall discuss them briefly.

Years in School. It might seem that the effect of any one school on the growth of its pupils ought to be proportional to the length of time the pupils have been associated with it. It is exceptional for us to find a correlation, however, between years of attendance and conduct tendencies. Such influence as the school may exert is brought to bear, apparently, in such a way as to give little significance to the extent of association. We have reported these correlations for honesty and service. In the case of persistence they are .189, .282, and .116 for X, Y, and Z. Since persistence correlates also with age and since there is a slight association between age and length of attendance in X and Y, a part of this connection between persistence and attendance in these schools may be due to the common factor of age. For children all of the same age, the r's just quoted would be .170, .092, and .129. Thus in Y, where there is considerable association between length of attendance and service, we find practically no association at all between length of attendance and persistence.

The case for the total inhibition score is similar. The r's with length of attendance are $-$ .041, .108, and $-$ .003 for X, Y, and Z. Whatever general advantage one population may have over

another, it is exercised from the very beginning of a pupil's career in the community in which he is now living.

This is well brought out by Figure 96, which gives the median persistence total for pupils who have just entered the school in which they now are placed, those who have been there two years, three years, etc. The effect of age may be readily observed among those in attendance 8, 9, and 10 years, who could not have been less than thirteen or fourteen years old. That is, there are *no* younger children among these, but there are some older children among those who have spent less time in any one school. Among these older children, as may be checked by glancing at Figure 75, population Z is not superior to X or Y. But those who have been a shorter time in attendance are conspicuously more persistent than the same groups in X or Y.

School Differences. It has already been pointed out that the superiority of Z in total persistence score is due to its score on the two puzzle tests. On the other three the Z pupils are inferior. On the two puzzle tests there was a time limit, and a study of the distributions reveals that more of the children of Z tended to use up the time on one puzzle than in the case of X or Y. If time had permitted, these more persistent children could have gone on working and might have brought the average of Z still higher. On the Magic Square test it was the lower grades which showed this tendency to use up the time available, whereas on the Ring and Cross test it was those in the upper grades who did so. This evened matters up between them and explains why there is only a low correlation between persistence and age in Z and why the older children were about the same in all populations on the puzzle score. To be sure, not many actually used up all the time; but it would be natural to expect that even the most persistent would feel increasingly the desire to get a look at the alternate puzzle before time was up.

The only problem, then, is to explain the tendency of the younger children of Z to work longer at these tests than the younger children of X and Y. We suggest as a possible answer that in conduct of this sort the younger children of the unfavored population might

well be expected to behave more nearly like the older children. This is partly checked by considering the association between persistence and age for children who have been the same length

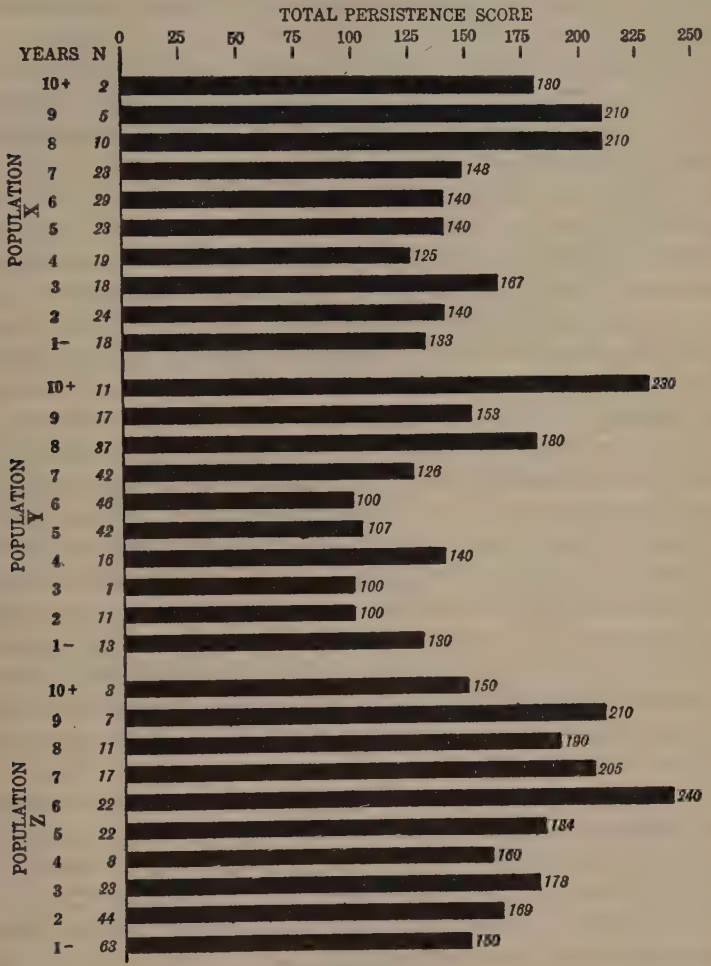


FIGURE 96

MEDIAN PERSISTENCE SCORES ACCORDING TO LENGTH OF ATTENDANCE AT EACH SCHOOL — POPULATIONS X, Y, AND Z

of time in one community. These partial *r*'s are .383 for X, .360 for Y, and only .178 for Z. We interpret this to signify the existence of a tendency in population Z for the younger children to behave with relatively greater maturity on the puzzle tests than is the case in the more favored communities X and Y.

We still have the fact of correlation of persistence with age and with grade in X and Y, as against the lack of correlation between persistence and length of attendance. It would appear that in conduct of this sort the general experience of the children is of greater significance than the school experience. But let us turn to the related problem of classroom morale.

Classroom Morale. Whatever may be the cause of the changes in the pupils' behavior from grade to grade, it is exhibited in the classroom among classmates and reflects the changing morale of the group. As we have shown in Book One and also in our *Studies in Deceit*, there seems to be evidence that the group morale is cumulative, tending to move in one direction. Where the influence of the school and community is helpful, quite apart from the length of time the individuals are in the community or in the school system, the morale of classes improves; but where the influence of the community is unwholesome, classes tend as units to get worse, quite apart from the length of time any given individuals are under this influence. Each pupil as he enters a class tends to become like the pattern of his class group, and a sufficient nucleus is carried on from year to year to determine the drift of attitude characterizing the class. This drift is the group's response to its environment rather than the independent responses of the individuals composing the group.

Some classes may be expected to behave differently from others, partly because of the teacher's influence and partly because back of each group is a history peculiar to itself. To separate the variations in classes, we present classroom means in Tables CIII and CIV of Appendix C. Parallel groups of the same grade in the same school would score approximately the same if there were no such specific group differences as those to which we have just referred. Comparison of the classrooms within grade 5 and within grade 6

in population Y, and of the classrooms within grade 8 in population Z, shows that these groups actually differ considerably from one another both in conduct and reputation. Class P, for instance, was far more persistent than class O (both of grade 5 in school Y). In reputation for persistence, the same group was also significantly higher than the parallel group of the same grade.* Class R was also lower than class Q (both of grade 6 in school Y), and class F was higher than class G (both of grade 8 in school Z).† Although the classes compared might *a priori* have been expected to score much the same since they were samples drawn from the same total grade population in the same community, they really showed large differences. With respect to the types of behavior incorporated in the tests, something had happened to them to make one class as a group excel and the other fall to an inferior position.

Idiosyncrasies of groups such as these are worth studying. For example, the fifth-grade class in population Z proved far more coöperative on the various service tests than the other classes and also scored second highest on the Persistence for Class test — it being, indeed, the only class scoring higher for class than for self — but it was low in the Persistence for Self test and was the very lowest in the inhibition tests. Evidently this class was more sensitive than the rest to appeals for help, as shown in service to one another and also to children in need with whom they were not acquainted. It was superior only in helpfulness, but in this type of behavior it was uniformly high.

In Chapter XI we mentioned a class in population Y which suffered from the fact that its teacher was ill and that it had not developed a satisfactory class spirit. This was class R of grade 6, referred to in the second paragraph preceding as having a low persistence score. Apparently this breakdown of the class affected individuals differently. Some seemed to lose morale more than others, for, as we pointed out in Chapter XI, the distribution on the Efficiency Coöperation test tended to be bimodal. That is, there was a tendency for the class to divide into two camps, one coöpera-

* The difference in test score is 2.9 sigma and in reputation score 5.3 sigma.

† These differences are respectively 3.0 and 2.6 times their SE's.

tive and the other uncoöperative. Its total service score, like its total persistence score, was low, and it shared with one other class the lowest score in inhibition. In contradistinction from the class in Z which proved distinctly *superior* in a single type of conduct, class R may show the results of a general break-up of morale as contrasted with the effect of some single constructive influence.

Another indication of classroom homogeneity is found in the resemblance of boys and girls belonging to the same rooms. The technique for measuring this was described in Chapter XI, page 214. Briefly, the tendency of a class to be different from other classes because of some factor or factors which are peculiar to itself, such as the teacher or its own history as a class, is shown by the correlation between the average score of boys and the average score of girls. The bigger the correlation, the greater the effect of influences tending to make the classes into integrated units each having its own peculiar character. In the case of persistence and inhibition we find about the same degree of classroom homogeneity, as thus measured, as we did in the case of service tendencies, with a somewhat more notable distinction between classes in respect to persistence than in respect to inhibition.*

These evidences are suggestive only, however, and are not precise measures of the extent to which classrooms differ from one another.

* TABLE LXI

CORRELATIONS † BETWEEN SELF-CONTROL MEDIANS OF BOYS AND GIRLS — ALL CLASSES IN POPULATIONS X, Y, AND Z

	POPULATION			
	X	Y	Z	XYZ
Total persistence score	.771	.867	.509	.736
Total inhibition score	— .100	.621	.634	.443

† The coefficients for the separate populations are rank difference *rho*'s, and for the total, product-moment *r*'s.

CONCLUSIONS CONCERNING HOME AND SCHOOL IN RELATION TO SELF-CONTROL

Brothers and sisters resemble one another in persistence to about the same extent as in service tendencies. The *r*'s run from .136 to .458 for various tests and combinations of tests. In the case of inhibition, they run from .189 to .344. The average of the *r*'s for the separate inhibition tests is .265 and for the separate persistence tests .212. Such resemblance as is here indicated cannot be accounted for by the influence of the home background factors which we have measured nor by the resemblance of sibs in intelligence.

The more intelligent parents in X and Z have children whose persistence and inhibition scores tend to be slightly higher than in the case of children of less intelligent parents. In population Z the correlation of parental intelligence with the persistence scores of children is .301 and with inhibition scores, .407.

The influence of schools has no relation to the length of time a child is associated with any one school. The *r*'s for X, Y, and Z are .189, .282, and .116 for persistence and − .041, .108, and − .003 for inhibition.

In all populations, inhibition scores increase with grade, and in X and Y the same is true in the case of persistence; but in Z the older grades are less persistent. Z as a whole has higher scores, but these are due to the superiority of the younger children over the children of corresponding ages in X and Y, for the older children are less persistent in Z than in X and Y. These contrasts might be due to some failure of Z to continue to stimulate the pupils in grades 7 and 8, or they might be due to something in the out-of-school environment which makes the younger children of Z behave more as the older children do. The latter is the more reasonable explanation, since the Z type of pupil is superior in certain forms of persistence even in X and Y.

The influence of the classroom is demonstrated in a variety of ways, and it seems clear that each room develops its characteristic morale, which accumulates or builds up from year to year.

CHAPTER XXVII

OUT-OF-SCHOOL EXPERIENCE AS A FACTOR IN SELF-CONTROL

It has by this time become apparent that self-control is not only highly specific but is also the result of miscellaneous causes, no one of which seems to stand out as a prepotent factor. The strongest associations so far noted have been those of age, intelligence, sex, school experience, and such aspects of environment as are indicated by community, nationality, and the home; and these have been bound up less closely with our inhibition scores than with persistence. It remains now to bring together such facts as are available regarding the relation between self-control and Sunday-school attendance, experience in clubs and summer camps, and attendance at motion picture shows.

SUNDAY SCHOOLS

Sunday-School Attendance and Persistence

Population Y. As we pointed out in our discussion of service, we have for population Y the actual record of attendance in the case of three religious groups, the Catholic, Dutch Reformed, and Baptist, including some 84 children. For the rest of Y we know the fact but not the amount of attendance.

Length of Attendance. The correlation between length of attendance and the persistence total, with age constant, is .17. Graphically the facts, which do not include the Catholic group, appear in Figure 97. There is a fairly steady rise for attendance from one year to six years, and a falling off for those who have been in Sunday school seven years or more.

It is interesting to note from Figure 97 that the 12 who are not enrolled anywhere have a median score of 168, which is 23 points above the median for those who are enrolled.

Regularity of Attendance. For this we have 72 records, and the median scores for those who attend once, twice, etc. a month, on the average, are shown also in Figure 97. Again no steady change is noticeable, but those who are *moderate* in regularity have the highest scores.

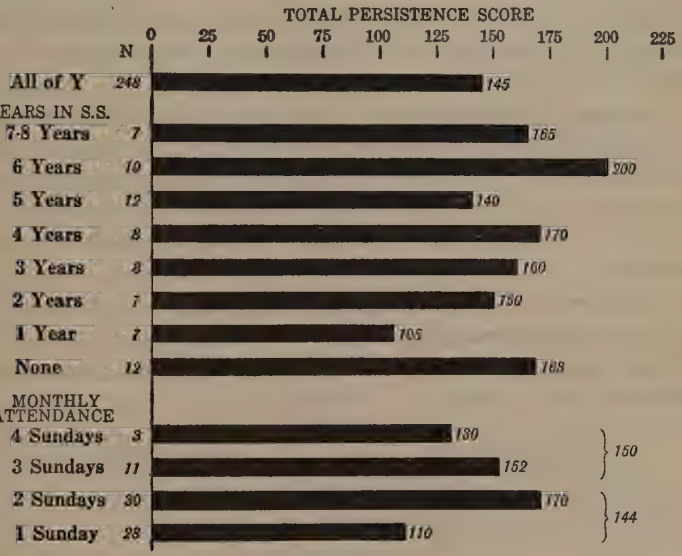


FIGURE 97

SUNDAY-SCHOOL ATTENDANCE IN RELATION TO MEDIAN PERSISTENCE SCORES FOR SELECTED GROUPS IN POPULATION Y

Populations X and Z. For these populations we have only the pupils' own statements regarding their attendance, and these probably cannot be taken at their face value. The correlations, however, between years of attendance and persistence, with age constant, are for X and Z respectively .241 and .087. There is a small general trend upwards in scores as years of attendance increase. But the 50 children of Z who, according to their own

statement, have never been enrolled in Sunday school have persistence scores well above the average for Z and 17 points above the median of those who attend Sunday school. These 50, however, are largely of the groups which on the whole have higher persistence scores than the average anyway, so that too much stress must not be laid on the fact that some of them do not attend Sunday school. In the case of X, those not enrolled also have a high average, but the 10 cases involved consist of 6 Jews, 3 Catholics, and 1 Protestant. The Jews probably go to some religious school during the week while the Catholics as a group have high scores anyway

In the records of regularity of attendance, we have the ambiguity noted in Book One, namely, that the pupils tended rather generally to report themselves as going to Sunday school four times a month. When we compare the persistence scores of those who say they go once or twice a month with the scores of those who say they go three or four times a month, we find that it is the irregular who are the more persistent. The differences are small, however, and must not be taken too seriously; otherwise it might have been inferred that the phrase, " He persists in staying away," had some general significance as a description of character.

Sunday-School Attendance and Inhibition

We shall report the facts for only the sum of the Story Inhibition test and the two manipulation tests.

Population Y. The correlation between years of attendance and inhibition is .079. Those who have stayed longest with the Sunday school have the highest scores, with the newcomers presenting a close second. Those not attending at all have a median of 14.0, which is a point above the average for Y. The medians, beginning with those not attending at all and for each additional year of attendance, are: 14.0, 14.7, 13, 13, 13.5, 12, 15.3, 15.2.

With regard to regularity of attendance, those attending two or three Sundays in four have the advantage over those who are either very regular or very irregular. Beginning with those attending once a month, the four medians are: 12.9, 13.5, 15.3, 12.5.

If the two more regular groups were combined, they would show slightly higher inhibition scores than the two less regular groups.

Populations X and Z. The correlation between inhibition and length of attendance at Sunday school, taking the pupils' statements with a grain of salt, is − .004 for X and − .005 for Z. No clear-cut differences appear to which any significance can be attached.

In the matter of regularity, the children who report that they attend Sunday school three or four times a month are *better* inhibited in X (12.4 against 10.5) and *less* inhibited in Z (11.6 against 15.2) than those who report an attendance of once or twice a month. But the numbers are few, and the data are given for interest rather than for information.

Conclusions concerning the Relation between Sunday-School Attendance and Self-Control

There is a slightly stronger association between persistence and length of Sunday-school attendance than between persistence and day-school attendance. As in the case of day schools, there is no relationship between inhibition and attendance. No constant relation is found between regularity of attendance and self-control. When due allowance is made for the constituency of the attendant and nonattendant groups, such differences as appear can well be accounted for on other grounds than the presence or absence of Sunday-school influence.

SYSTEM Y

All organizations for boys and girls make much of the virtues of persistence and inhibition. To stick to a task until finished and to resist distractions and refrain from the indulgence of self at the expense of others have always been regarded by western peoples as characteristic of the ideal man, whether he be clothed in the garments of the savage, the knight, the pioneer, or the woodsman. And now that girls are receiving much the same training as boys,

these fundamental traits have not been neglected by programs planned for them.

In view of the widespread emphasis on these types of conduct, it might be expected that a brief experience in some organization would not be likely to raise its members above the general level. Nevertheless, it might well be anticipated that the very emphasis placed on self-control by schemes such as system Y would tend to attract children who for one reason or another are already imbued with this ideal. Figures 98 and 99 show that any such superiority among the members of these organizations as may exist is certainly not general, and in fact does not appear in our tests.

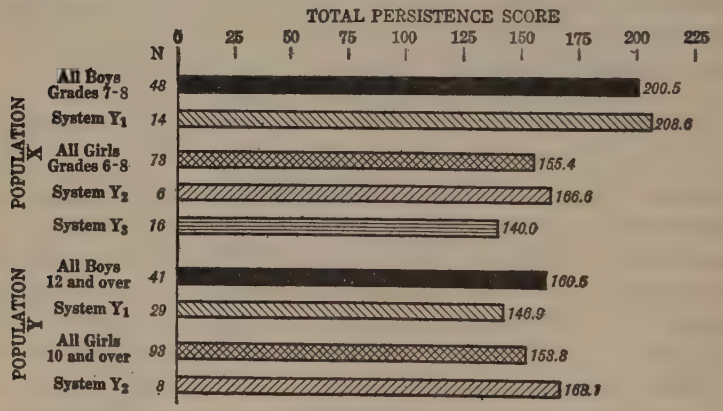


FIGURE 98

MEAN PERSISTENCE SCORES OF MEMBERS AND NON-MEMBERS OF ORGANIZATIONS — POPULATIONS X AND Y

System Y and Persistence. Taking the persistence total for our first comparisons, we find from Figure 98 that the members of system Y_1 score slightly higher than the average of boys of grades 7 to 8 in population X and slightly lower than the average of boys over 12 years of age in population Y. These differences are all well within the limits of chance.*

* All the members in population X were in grades 7 and 8. The difference is .4 σ in favor of the members. In Y the comparison is made with boys over

Eight boys in one class of Z had a median persistence score of 220 as against 206 for all the boys of this class.

As for Y_2 girls, those of population X are slightly more persistent than all girls of the grades involved, and in population Y the facts are exactly parallel.*

Y_3 appears only in population X. Members are less persistent than the average for the girls of their own classes.†

These differences would all have been slightly greater if the members had been compared with non-members rather than with the whole population, but the differences and the numbers involved were all so small that it did not seem worth while to take extra time for making the more refined computation.

System Y and Inhibition. Figure 99 shows what we have found. In population X the boys who belong to Y_1 are slightly less inhibited. They would be slightly inferior probably 7 times in 10 similar experiments.‡ In Y the difference is relatively greater.‡ It would be in favor of the non-members about 9 times in 10. In Z the children belonging to Y_1 are superior, but again the difference is barely perceptible.‡

As for Y_2 girls, in population X the difference in their favor is fairly substantial ‡ and could be expected to continue to be in their favor 986 times in 1000 similar studies. But in Y the difference is slightly against them. Likewise in population X ‡ the members of Y_3 are on the average slightly less inhibited than the girls as a whole.

Considering only the Stories, Puzzles, and Safes, we find no differences between the girls of Y_2 or Y_3 in population X or Y, but

---

12, since members must be at least 12. For population Y the difference is about .7 σ in favor of those who are not members.

It should also be noted that the members of Y_1 have higher cultural scores than non-members. The differences in these scores are .7 σ, 2.9 σ, and 2.5 σ, respectively, for populations X, Y, and Z.

* In X the difference is .41 σ and in Y .46 σ in favor of members.

† The difference is 1.0 σ in favor of the whole group.

‡ Population X: boys minus Y_1 = .53 σ; Y_2 minus girls = 2.2 σ; girls minus Y_3 = 1.0 σ. Population Y: boys minus Y_1 = 1.2 σ; girls minus Y_2 = .3 σ. Population Z: Y_1 minus boys = .27 σ.

the boys of system Y_1 have slightly lower scores than the rest in both populations X and Y. Nine boys in one class in Z who are members of Y_2 have a median of 14.7 as against one of 15.3 for all the boys of this class.

Conclusions concerning Clubs and Self-Control

Members of the three clubs included in this study show only chance differences from the bulk of the population of the same sex

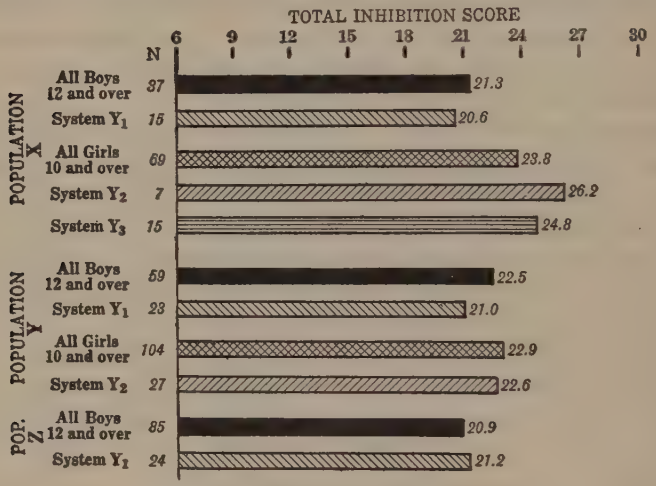


FIGURE 99

Mean Inhibition Scores of Members and Non-Members of Organizations — Populations X, Y, and Z

and age in the types of behavior included in our tests of persistence and inhibition. In all, eight comparisons were made between boys of System Y and all boys of the same population; only two of these show the members to excel, and one of these two is obviously a chance difference. We have nine comparisons in the case of the girls. Five of these are favorable to members, three are unfavorable, and in one instance the average of members is exactly that of non-members. But again these differences, with one possible exception in the case of the superiority of Y_2 members of

population X, are such as might be expected from the operation of chance. It is apparent, therefore, that the members of these clubs are, on the whole, just like other boys and girls of their own age and sex in the types of conduct which we have called self-control.

SUMMER CAMPS

The 42 children of population X who had had the privilege of attending summer camps were just about average in both the persistence and inhibition scores.

MOTION PICTURES

We came upon some interesting contrasts in our study of service tendencies between those indulging freely and rarely in motion pictures. In the case of persistence and inhibition, we have worked out similar comparisons for the total scores only, and the results are presented in Figures 100 and 101.

Motion Pic ures and Persistence. The median for each degree of attendance is shown for each population in Figure 100. No clear distinctions appear save in population Y between the group which attends now and then, but less than once a week, and all the rest, and in Z between those who attend fewer than four times and those who attend four times or more. The children who report themselves as never going at all in Y have very low scores. In general, those who go often to motion picture shows have lower persistence scores.

Motion Pictures and Inhibition. The relation between attendance at motion pictures and any form of conduct is partly a function of the type of population involved. Figure 101 brings out these contrasts so far as the total inhibition score is concerned. It may be seen that in X it is those who do not go at all who show the higher scores, and from that point there is a steady decline until we come to those who go four or five times a week, and here the medians tend to rise once more. Not one of the eleven who do not

go at all scored lower than 20 (23.2 is the median for X) on the inhibition tests, whereas of all the rest 26% scored below 20.*

In population Y the facts are somewhat different. Just as in the case of persistence, it is those who go occasionally, but less frequently than once a week, who make the best showing.† Taking as one group those going three or more times a week and as another

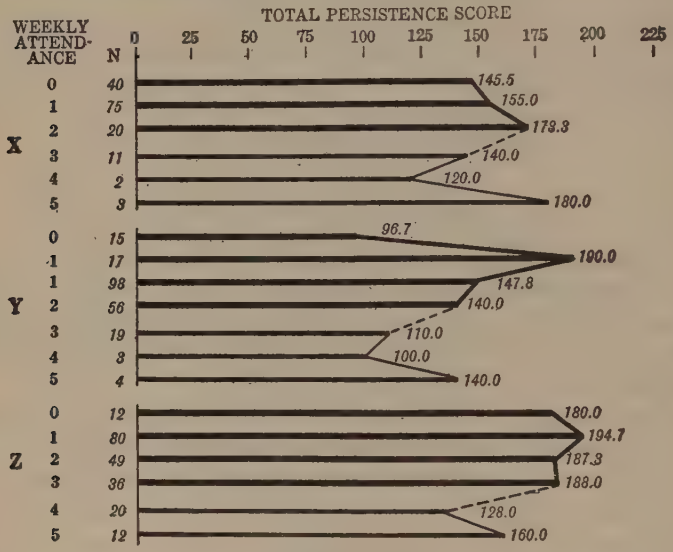


FIGURE 100

MOVIE ATTENDANCE IN RELATION TO TOTAL PERSISTENCE MEDIANS — POPULATIONS X, Y, AND Z

group those going fewer than three times a week, we find a decided difference in favor of those who attend less often.

In population Z there is a steady decline in score as attendance

* The difference between the non-goers and the others is 2.6 times its SE. The difference between the non-goers and those going less than once a week is 1.9 σ.

† The difference between this group and all the rest is 1.9 times its SE, and the difference between this group and those attending not at all is 2.2 σ in favor of the former.

increases. Indeed the r between frequency of attendance and inhibition is $-.231$. The difference between those attending more

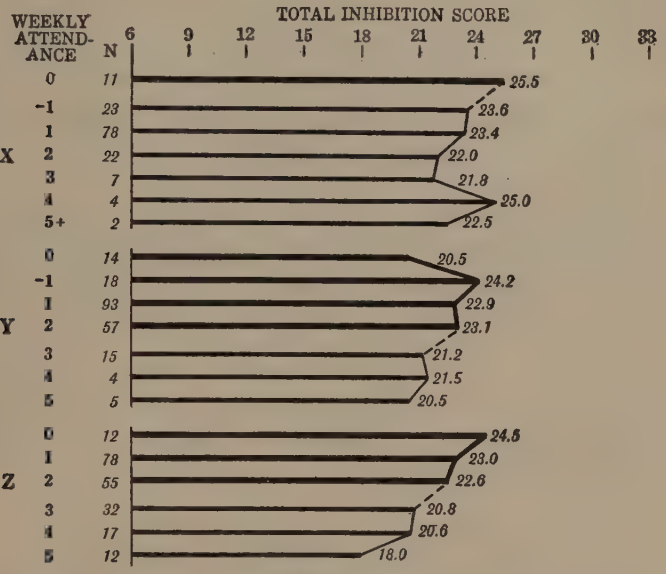


FIGURE 101

MOVIE ATTENDANCE IN RELATION TO TOTAL INHIBITION MEDIANS — POPULATIONS X, Y, AND Z

than twice a week and those attending fewer than three times a week is striking.*

CONCLUSIONS CONCERNING MOTION PICTURE ATTENDANCE AND SELF-CONTROL

Those who attend motion pictures less frequently than their schoolmates do are inclined to exert greater effort and greater self-restraint. It is a question whether this results as a direct effect of the pictures themselves or from characteristic ways in which habits of attendance on motion pictures develop among different sections of the population of a given community.

* It is 2.5 times its SE.

CHAPTER XXVIII

MISCELLANEOUS FACTORS ASSOCIATED WITH SELF-CONTROL

In addition to the concomitants of self-control that we have so far discussed there remain a few facts which are not readily classified with the material already covered and which we may therefore bring together in one place. These facts are suggestibility, cleanliness, sociability, friendship, and speed ability. We shall also report the effect of being in the majority and shall compare the well controlled with the poorly controlled in various respects. Of the specific nature of self-control not much more need be said, but we shall devote a paragraph to this problem also.

SUGGESTIBILITY AND SELF-CONTROL

The tests for measuring suggestibility have been sufficiently described heretofore,* and we reported in Book One that the ability to resist suggestions of the type presented in the test is only slightly associated with tendencies to be of service, although strongly associated with honesty. The facts for self-control are about the same as for service — the association is positive but slight.

Suggestibility and Persistence. In the case of certain persistence tests, as for example the Cross and Ring, one determining factor in the length of time a child works may be the length of time his classmates work. Such social stimulation had a favorable opportunity to operate in this test, since as soon as they gave up the children were expected to go to the examiner and get the other puzzle to try. If a child decided to quit after he had seen a few

* The Otis test. See pages 233 to 234 and also *Studies in Deceit*, Book One, pp. 277 ff.

of his friends quit, he ought not to have had the same score as the child who quit last. The only way in which we have allowed for this is by recording the length of time the pupils worked at the first puzzle, and no account is taken of the fact that classrooms differ greatly in the way the pupils behaved on this test. In some they quit in bunches. In others there is a more even distribution through the period. We shall speak presently of this phenomenon.

Our only measure of suggestibility is the Otis test, which does not pretend to measure social suggestion of the sort we have been discussing. Even so, however, the correlation between the ability to resist the suggestions on this test and the ability to score on the persistence tests is .283 for X, .231 for Y, and .118 for Z. The median persistence scores for six levels of suggestibility in population X are: 100 137, 144, 172, 183 150

Before concluding that persistence and resistance to suggestion are directly related, we should consider whether age is a determining factor in both. As the correlations between age and resistance to suggestion are not far from zero in our populations, age is obviously not involved. But how about intelligence? This correlates with both persistence and resistance to suggestion. In fact, when this common factor is allowed for the partial correlation between persistence and the ability to resist suggestion drops to .192 in X, .152 in Y, and .085 in Z. Thus in X and Y a very slight association remains.

Suggestibility and Inhibition. There is less indication of relationship here than in the case of persistence. The r's are $-$ .060 for X, .079 for Y, and .192 for Z. When we omit the Picture test and retain the Stories, Puzzles, and Safes, on which the pupils could tell easily what was going on in the room, the r's become .035, .102, and .242. The r for Z (.242) is evidently partly a function of intelligence, for in this population intelligence and the three inhibition tests are correlated to the extent of .374. Here the partial r between these three tests and resistance to suggestion, with intelligence constant, is .074.

Social Suggestion and Self-Control. Another way of determining the extent to which the self-control and persistence scores were

influenced by the actions of classmates is by examining the distributions themselves. The two tests that yield most readily to this type of treatment are the Cross and Ring and the Magic Square. On these two tests the score is the length of time the pupil worked on the task before giving up. Furthermore, the way in which these two tests were administered makes them especially susceptible to social suggestion. It will be recalled that the children choose a puzzle, work on it until they wish to quit, and then *take it up to the examiner* and ask for another one. The fact that a child leaves his seat and walks forward is a strong suggestion to others to do likewise. If there was a tendency for others to quit upon seeing one of their classmates do so, one would expect the distribution of time scores to be bunched. Suppose that at the end of the first five minutes one pupil goes forward for another puzzle. If this acts as a suggestion to others, then we would expect to find several going up at the same time. This would bunch the time scores.*

The distribution curves for both the Cross and Ring and Magic Square tests are shown for eight classes in Figures 102 and 103. The solid lines represent the Cross and Ring, and the dotted lines represent the Magic Square. In some groups we see strong indications of bunching. For example, in group A no one quit on the Cross and Ring until after 10 minutes; then one gave up, and by the end of 20 minutes 15 more had quit and by the end of 25 minutes 9 more. Then there was a lull in the tide, which rose again at 30–35 minutes, when 11 more stopped working, and then they straggled in till the end of the period. It is interesting to note that in this group the curve for the Magic Square follows almost exactly that for the Cross and Ring.

Group B presents a different picture altogether. On the Cross and Ring test one quit in the 15–20 minute interval, 2 more by the end of 25, 4 more by the end of 30 minutes. Evidently the first one who gave up did not start a stampede. In the ten-minute interval between 40 and 50, 14 went up for another puzzle; but

*Furthermore, the examiner reminded the pupils every five minutes that they might exchange puzzles whenever they wished.

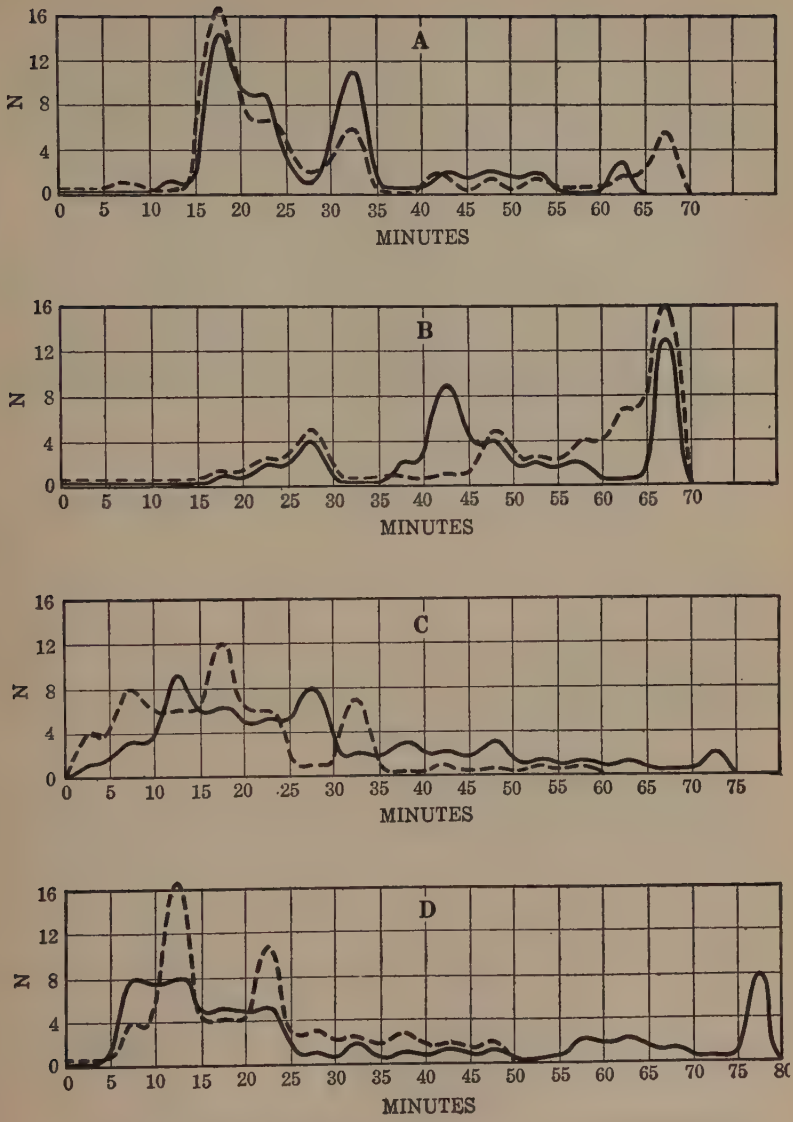


FIGURE 102

EFFECT OF SOCIAL SUGGESTION ON PERSISTENCE OF PUPILS

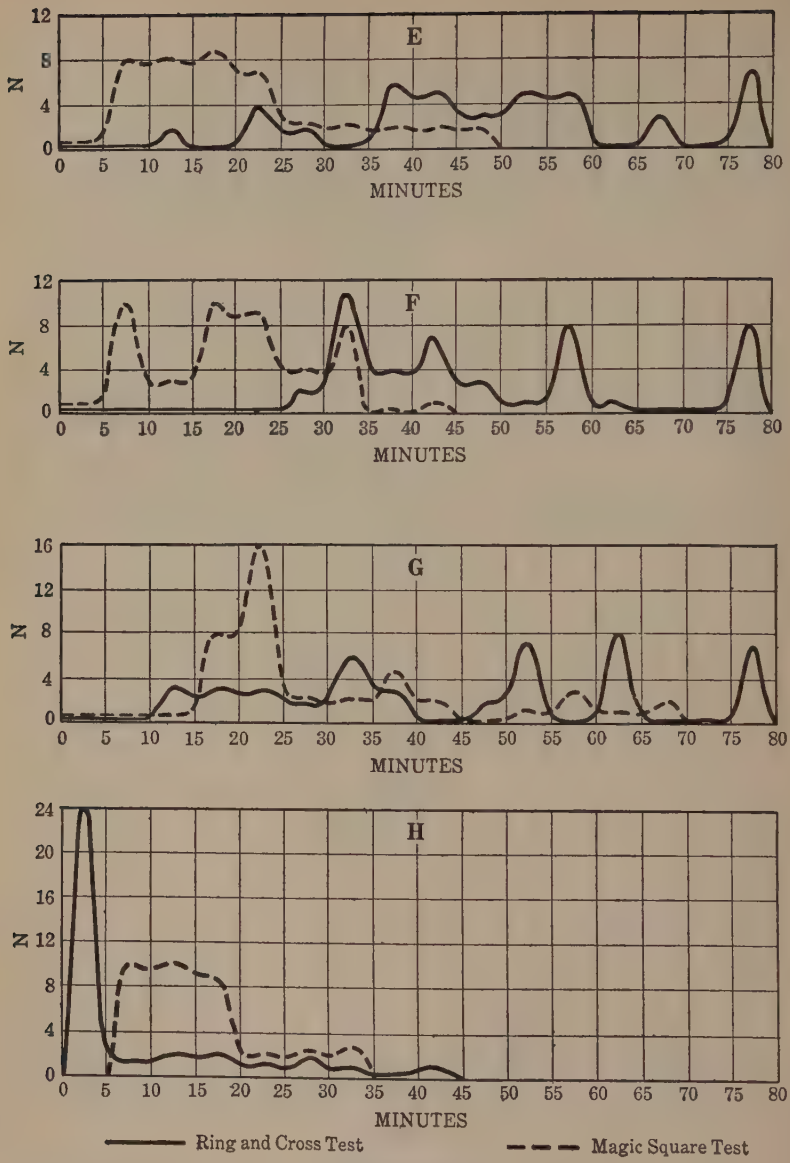


FIGURE 103

EFFECT OF SOCIAL SUGGESTION ON PERSISTENCE OF PUPILS (CONTINUED)

at the end of the period 13 were still going strong. This is one of the few groups in which many were still working at the end of 70 minutes.

Group H is a queer one. In this group there were 34 pupils, 24 of whom quit before the first 5 minutes were up. Evidently the first one to quit started the stampede. Here we have a clear case of social suggestion. There are two or three other groups (not plotted) that behaved in somewhat the same fashion as H.

The conclusion is that the matter of social suggestion varies from group to group. In at least 4 out of the 22 groups tested there is evidence of a considerable amount of it. Group H is of this type. There are ten or twelve other groups in which the evidence is the other way, of which groups B, E, and G are typical. Then there is a middle type, in which it is evident that some pupils are very susceptible to this influence and others are not; of this type A, C, and D are good illustrations.

A more detailed examination of these time scores would no doubt reveal some interesting data bearing on the problem of social stimulation.

Conclusions. The type of suggestibility measured by the Otis test is not generally related to persistence and inhibition, save when the factor of intelligence is ignored, in which case the scores on the Otis test show a low positive association with persistence in all populations and with inhibition in population Z. The factor of social suggestion, however, as indicated by the specific behavior of different classrooms, does seem to be influential in determining the way the pupils behave in some of our tests and possibly on others. A more detailed study of the Otis scores of classes showing a high degree of social suggestibility might exhibit a closer association between self-control and this test than is found for all classes combined. Time has not permitted us to go further into the matter.

CLEANLINESS

This score, it will be recalled, was secured by the examiners of the American Child Health Association in their study of a large

number of pupils of population Y. Dividing the cleanliness scale somewhat arbitrarily into sections ranging from the extremely clean, or 0's, to the moderately clean, scoring from 2 to 9, the rather unclean, scoring from 10 to 17, and the extremely unclean, scoring from 18 to 21, we find some interesting contrasts, as shown in Figure 104.

In the case of persistence, the difference is quite substantial between the very clean (0's) and the rest. The former would have

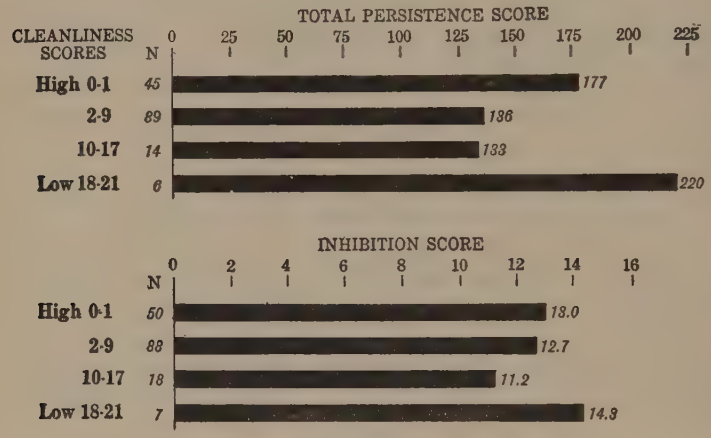


FIGURE 104

MEDIAN SELF-CONTROL SCORES FOR DIFFERENT DEGREES OF CLEANLINESS — POPULATION Y

the advantage, probably, in 994 out of 1000 similar experiments. The differences in inhibition, however, are not so great* even though there seems to be a fairly steady rise in scores as cleanliness increases — except for the extremely dirty, who are both better inhibited (lazy?) and more persistent (obstinate?) than the rest.

* The extremely clean are 1 σ better inhibited than the rest.

SOCIABILITY AND SELF-CONTROL

We measured sociability by counting the number of times a pupil was mentioned as a friend or companion by his classmates or others in the school. The correlations with this evidence of being able to make friends and the conduct tendencies handled by our tests of persistence and inhibition are all zero or nearly zero. But it is a curious thing that, as in the case of service tendencies, the most popular or sociable children tend somewhat to score nearer the middle of the conduct scales than at either extreme. In population X 43 children have persistence scores above 200, but only one of these was named more than four times as a friend. On the other hand, 21 were named more than four times as a friend and only one of these has a persistence score of more than 200. Subdividing the whole population of X in this way, we have the following fourfold table:

Sociability \ Persistence	Scoring less than 200	Scoring 200 or more
Having more than four friends	11.2%	.6%
Having four or fewer friends	64.2%	24%

This shows the per cent falling in each section and illustrates the fact that the most persistent do not have the most friends. Two-thirds of the 21 named by more than four schoolmates have persistence scores below the average of the school. The facts are somewhat similar for populations Y and Z.

In the case of inhibition, considering only the Stories, Puzzles, and Safes, all we can say is that those who are regarded as friends by a good many others do not have *low* inhibition scores, which reverses the facts for persistence, where we said that this group does not have *high* persistence scores.

FRIENDSHIP AND SELF-CONTROL

A somewhat different approach to the problem of the influence of sociability takes account not only of the number of friends but also of their association. It would be highly desirable to supplement such studies as we have made by direct observation of the nature and extent of such association, but we are able to report only its effects so far as these appear in the resemblance of friends. Using the technique already described in Book One and also in *Studies in Deceit*, we paired off each pupil with someone named by him as a friend, regarding the pair as an association unit with two scores to be correlated. If the correlation is zero, then the members of the pair resemble each other no more than they would by chance. If it should be + 1.00, this would mean that the two members of each pair had the same score.

In order to throw some light on the extent to which classroom association influenced the resemblance of friends, we formed pairs from within the same classroom and pairs from different classes. The results are all given in the tables below and on page 437.* Comparison with Table XXXVI will show that friends tend to resemble one another about as much in persistence as in service,

*TABLE LXII

RESEMBLANCE OF FRIENDS OF SAME CLASSROOM IN SELF-CONTROL

TEST	POPULATION X		POPULATION Y		POPULATION Z	
	N	r	N	r	N	r
Cross and Ring plus Magic Square	186	.357	228	.309	200	.312
Persistence for Self plus Persistence for Class	194	.201	224	.272	242	.124
Total persistence	146	.417	204	.206	168	.293
Picture Inhibition	224	.004	246	− .269	248	.065
Puzzles, Safes, and Stories	154	.126	222	.124	200	.364
Total inhibition	134	.115	198	− .001	176	.223

but the resemblance in inhibition is considerably less. It is interesting to note that on the Cross and Ring and Magic Square test combination and the Puzzles, Safes, and Stories combination, where the conduct of one child is more or less conspicuous, the resemblance between friends of the same class is considerably greater than on the other tests, where what one child does cannot readily be seen. This might be attributed to the very natural tendency which a youngster might have to do what his friend does were it not for the fact that the same contrast appears in all populations when the friends are drawn from *different* classrooms and one member of the pair could not know, therefore, what the other was doing. It should be noted, however, that friends from different classrooms show far less resemblance in self-control than friends from the same classroom. That they show any at all is of considerable interest.

SPEED ABILITY AND SELF-CONTROL

In view of the fact that several of our tests are based on the sustained-speed abilities of pupils in mechanical tasks, it seemed desirable to determine the extent to which these abilities were of influence in determining the persistence and inhibition scores

TABLE LXIII

RESEMBLANCE OF FRIENDS OF DIFFERENT CLASSES IN SELF-CONTROL

TEST	POPULATION X		POPULATION Y		POPULATION Z	
	N	r	N	r	N	r
Cross and Ring plus Magic Square	54	.152	150	.176	78	.221
Persistence for Self plus Persistence for Class	52	− .370	148	− .016	108	− .017
Total persistence	46	.080	136	.200	64	− .107
Picture Inhibition	68	.152	174	.038	108	− .015
Puzzles, Safes, and Stories	52	.205	144	.134	82	.041
Total inhibition	46	.249	136	.085	74	− .091

derived from measures of them. We had an excellent though specialized measure of sustained speed on the Persistence for Self and Persistence for Class tests, which required fast work on simple additions for twenty-four minutes, each of which was separately scored. The total amount done in these twenty-four minutes was used as a sustained-speed score. This score correlates with the Picture Inhibition .252, .655, and .293 in populations X, Y, and Z.* This test uses simple additions also. As sustained-speed ability correlates with age .128, .324, and .304, it is apparent that the slight correlation of age with inhibition on this Picture test is a function of sustained-speed ability; and we suggested that ability to keep on adding rapidly is partly a function of the ability to resist distractions. The former increases with age but not the latter.

These correlations between sustained-speed ability and the ability to resist distractions should not be confused with correlations between *initial* speed ability and inhibition. When we take as our measure of speed ability the work done on the four minutes of practice *preceding* the Picture Inhibition test, we find that this correlates with resistance on the picture distractions only .119, .090, and .046 in X, Y, and Z. That is, those who do best on this preliminary test do not materially tend to do better than the rest when distraction is presented — here initial spurt of speed is apparently not a function of distractibility. But later on, when the fatiguing speed test is given, which lasts twenty-four minutes, the factor of distractibility enters into the picture, as the *r*'s just quoted show (.252, .655, and .293).

Sustained-speed ability, as measured by the twenty-four minute test, does not correlate in X or Y with the sum of the Puzzles, Safes, and Stories tests, which have no speed factor in them. The *r*'s are .078, .049, .246, and .020 for X, Y, Z, and all combined. Z exhibits some association even on these tests.

The two persistence tests which might be expected to show some correlation with speed ability are the Persistence for Class and

* If these *r*'s could be corrected for attenuation, they would be still higher.

Persistence for Self tests. These scores are found by taking the difference between the average of the first four minutes and the average of the last four minutes of the work done to win a prize for self, and the same measure in the case of work done to help the class win a prize. The speed ability represented in the total amount done in the whole twenty-four minutes correlates with Persistence for Self .193 in X, .185 in Y, .178 in Z, and .061 in all combined, and with Persistence for Class .034 in X, $-.175$ in Y, $-.050$ in Z, and .074 in all combined. On the total persistence score the corresponding r's are .186, .302, .144, and .336. But this relation between the total persistence score and speed ability is reduced when the association between age and speed ability is allowed for.* A slight association between persistence and speed in X and Y is still apparent, however, even for children of the same age.

THE EFFECT OF MAJORITIES ON SELF-CONTROL

We have already pointed out that classrooms differ in persistence and inhibition. It is also true that the classes themselves are not always homogeneous groups. This is particularly true as regards sex. In some groups the boys are superior, and in others the girls are superior. In our studies of deception and of service we have drawn attention to the fact that the majority sex in population Z exhibits the more aggressive behavior. The question now is whether this is true also in the case of our present scores.

In working this out we followed the practice described in Chapter XIII. The ratio of the number of boys to the number of pupils in the whole class was correlated with the ratio of the median of the boys to the median of the girls. In population Z the resulting correlations are .85 for persistence and .30 for inhibition.† In population Y they are about zero. The corresponding r reported for service was .70 in Z and zero for Y. In *Studies in Deceit*

* The partial r's, with age constant, are .149, .188, and .099.

† In population X we have two separate schools, dividing between an elementary and a junior high school. The corresponding r for this situation is about zero.

we mentioned a comparable r of .51 for dishonesty. This seems to indicate that to be of the majority sex has in at least one population an apparent influence on behavior. The association is strong enough in the case of inhibition to lead us either to question the soundness of the judgment previously expressed, that it was the aggressive behavior which was influenced by being in the majority, or else to conclude that inhibition is not passive do-nothing-ness but may indeed represent a more truly active response to a situation than uncontrolled yielding to the temptation of the moment. In the Picture Inhibition test, of course, the inhibited behavior was actually the aggressive behavior, involving great concentration and effort in doing the examples.

CHARACTERISTICS OF CONTRASTING GROUPS

One of our problems has been to throw some light on the relative amount of association between self-control and such facts as age, intelligence, and home background. In Chapter XIV we supplemented the usual method of correlation by a comparison of the upper and lower quartiles (that is, the 25% most coöperative and the 25% least coöperative) with respect to a large number of the concomitants we had been studying. This procedure materially aided us in forming some judgment as to whether differences in the tendency to help others is more intimately associated with differences in age than with differences in intelligence or culture or what not.

We have made similar comparisons of the most and least persistent children and the most and least inhibited. As before, we examined the actual distribution of each classroom and counted down from the top about one-fourth of the total number in the room and then counted up from the bottom about one-fourth, leaving about half the cases. Attention was paid to the position of the average so that the group left out of consideration would be what the statistician would call the interquartile range. This was done for boys and girls separately since the sexes differed considerably in most classes. The results show that with regard to

the concomitant facts studied the extremes in self-control are far less differentiated than the extremes in service.

Groups Contrasted in Persistence. The extent to which the least persistent group is like the most persistent group is indicated by the per cent of overlap. The smaller the overlap, the greater the difference. If the overlap is 50%, the two groups are approximately the same. If it is 0, then none of the members of one group reaches the median score of the other. The details are given in Table CV of Appendix C. There is considerable difference in populations as to the amount of overlap, but for all populations combined the per cent of overlap for each of the factors studied is as follows: health (Y only), 34.7; age, 36.5; reputation, 40.0; resistance to suggestion, 40.4; intelligence, 41.6; physical condition, 44.4; emotional stability, 44.6; cultural level, 50.0. Evidently the two groups are much alike.

The significance of the differences between the contrasted groups is, of course, as important as its amount. As might be expected from the small correlations reported, the probability that similar differences would occur again is not great enough to warrant our asserting any very strong association between persistence and anything else. These probabilities are also given in Table CVII of Appendix C. The most stable differences between upper and lower groups are in reputation, resistance to suggestion, absence of nail biting and other indications of emotional stability, and intelligence.

Groups Contrasted in Inhibition. The relatively well inhibited and poorly inhibited groups are more nearly alike than the relatively persistent and nonpersistent. In per cent of overlap the various factors studied appear in the following order for the three populations combined: reputation, 32.8; health, 41.2; emotional stability, 41.5; physical condition, 43.9; age, 44.7; cultural level, 45.5; intelligence, 46.8; and resistance to suggestion 53.9.*

* Strangely enough, the smallest overlap for any population is found in the case of our physical condition scores, although they show no general correlation with either persistence or inhibition. In population Y only 12% of the poorly inhibited equal or exceed the median physical condition score of the well inhibited. Here the significance of the difference between the extremes is

The probability that these factors differentiate the contrasted groups is not much beyond the limits of chance save for reputation, in which the groups could be expected always to continue to be different, and for intelligence and emotional stability, for each of which the probable expectation is about 98 times in 100.

THE SPECIFICITY OF SELF-CONTROL

The average intercorrelation of our four classroom deception tests was reported as .392 in *Studies in Deceit*. The corresponding figure for five service tests is .201, for five persistence tests .239, and for four inhibition tests .16. The last figure, as pointed out in Chapter XX, is probably so low on account of certain abnormal test conditions in populations X, Y, and Z. The figure for these tests in population C is .289. That types of conduct, to which we are accustomed to giving a class name, like honesty, coöperation, self-control, are congeries of specific behaviors is apparent. Indeed, as will be brought out in Volume III, these more general types of tendencies are about as closely related to one another as are the separate exhibitions of any one type. If our language customs permitted, we should learn to speak of persistences rather than of persistence, of inhibitive modes of conduct, not inhibition, and so on. Of all the tendencies studied, that which is represented by deceit is the most general; self-control comes next; and service, so far as we have tested it, is the most specific. We might be tempted to attribute the more generalized character of deceit to some such common factor as an ideal of honesty were it not for the fact that the conducts called helpfulness and coöperation are probably idealized to an even greater extent, and yet they are the least generalized. If the children have developed a general motive of charity or self-control, it obviously does not serve as a means of governing behavior. Confirmation of this inference will be found in our study of the relation of knowledge to conduct, which we shall report in Volume III.

---

nearly four times the SE of the difference. The overlap for all populations combined, however, is 43.9, as stated. See Tables CVI and CVIII of Appendix C.

PART III

Conclusions and Implications

CHAPTER XXIX

CONCLUSIONS AND IMPLICATIONS OF STUDIES IN SELF-CONTROL*

Book Two has presented sample studies in the measurement of certain conduct trends classified as persistence and inhibition. As in Book One and in *Studies in Deceit*, our purpose in making these studies has been twofold: first, to build a practical set of tests by which significant types of conduct could be measured and, second, while developing these tests, to find out all we could about the causes and nature of these tendencies. These two interests we have called our primary and secondary studies, respectively.

RESULTS OF PRIMARY STUDIES

Our measures of self-control have been developed in two series, the first called persistence tests and the second, inhibition tests. Both types involved a large amount of preliminary exploration and experiment, in the course of which many ideas were tried out. Although the results of only eleven techniques were finally used, we reported the details of several others for what they may be worth in further research. There follows a brief summary of the eleven techniques finally adopted:

---

*The conclusions and implications of the studies in service are given in Chapter XV.

PERSISTENCE

TECHNIQUE	NUMBER OF TESTS	NUMBER OF RESPONSES	TIME OF TESTING IN MINUTES	MOTIVE	RESISTANCE
Story Resistance .	3	No limit	30	Story interest	Difficult reading
Cross and Ring . .	1	No limit	70–80	Mechanical puzzle	Difficulty
Magic Square . .	1	No limit	70–80	Mental puzzle	Difficulty
Persistence for Self .	1	12	30	Individual prize	Fatigue
Persistence for Class	1	12	30	Class prize	Fatigue

INHIBITION

TECHNIQUE	NUMBER OF TESTS	NUMBER OF RESPONSES	TIME OF TESTING IN MINUTES	TEMPTATION	MOTIVE FOR INHIBITING
Story Inhibition* .	1	1	10	Know end of story	Depend on imagination
Story Inhibition .	1	2	15	Know end of story	Get a good mark by routine work
Puzzle Manipulation	1	7	Incidental	Handle puzzles	Not make work for others
Safe Manipulation .	1	6	Incidental	Handle safe and lock	Fair play
Picture Inhibition .	6	No limit	40	Look at pictures, etc.	Do your best
Ruggles Distraction*	1	No limit	20	Look at pictures, etc.	Do your best

* The results of these tests were not used, on account of errors in administration.

The battery of five persistence tests has a reliability of .89. There is little correspondence between the reputations we had available for persistence behavior and conduct on our tests. What was tested was not included to any great extent in the ratings. Some twenty-four more tests of the sort already devised would be needed to cover the ground with reasonable completeness.

The inhibition battery of four tests has a reliability of not less than .80, as measured by combining the four into two sub-batteries. The reputations for pupils in the matter of inhibition correspond more nearly to the types of experience represented in our tests than in the case of persistence; and if errors of measurement could be

eliminated, the opinions of teachers and pupils would agree with the test results to the extent of a correlation of .80. We believe that probably about twenty more tests similar to those reported would be needed for a complete measure of inhibition tendencies.

RESULTS OF SECONDARY STUDIES

The same body of material concerning age, intelligence, home background, and the like which was used in our study of service was available, of course, for the investigation of the concomitants of persistence and inhibition. The test program which was administered in the three contrasting populations, X, Y, and Z, included our entire set of tests, so that what we have to say about self-control applies to the same children as were dealt with in Book One.

Any attempt to summarize such complex relationships as we have been dealing with is bound to be inadequate and unsatisfactory. Two general conclusions are emerging which account for the intricate nature of our findings. The first is that the conduct trends and their relations to one another in individuals are the precipitates of specific experiences and are functions of the situations to which they have become attached by habit. The second is that these specific trends and relationships are gathered into patterns which represent not general ideas about conduct but, rather, specific group tendencies. That is, the group, with its accumulating experience, is a common factor in the situation to which its members are exposed.

The term " specific " is used as descriptive of a cross section of individual behavior. It represents, however, a useful theoretical reconstruction rather than direct observation. To be sure, observation does show positive correlations among the various types of conduct which common sense would classify as honesty, service, or what not; but the implications of our study are that the general law of transfer offers an adequate explanation of these correlations. The size of the correlation, of course, is a function of the extent to which the factors that determine a child's position on one test are identical with the factors producing his position on the correlated

test. Similarly "transfer," or the identity of response to two situations, is a function of the extent to which the potent factors of one situation are identical with the potent factors of the other. The bridge between these two observed phenomena — the correlation of two sets of responses and the identity of potent elements in two situations — is the theory of specificity, which interprets behavior as a function of the situation.

This point of view does not preclude at all the operation of ideals and standards but merely requires, first, that if they are to operate effectively they must be the prepotent factors common to the situations in which conduct is taking place and, second, that the operation of these ideational aspects of situations shall be demonstrated by the correlation of responses in these situations.

The phase of this problem which involves the functioning of groups as well as of individuals will be taken up more fully in Volume III.

The following statements will serve to remind the reader of the more outstanding conclusions reported in the previous pages.

1. The Relation of Self-Control to Age, School Status, and Intelligence. Brightness, age, and mental maturity are all involved in persistent behavior, but only slightly related to inhibition. (The respective *r*'s, given always for populations X, Y, and Z in this order, are: between persistence and age, .391, .438, and .169; and intelligence, .224, .226, and .091; between inhibition and age, .015, .134, and .136; and intelligence, .071, .003, and .269.)

In the unfavored community persistence is less in the upper grades than in the lower, but in X and Y there is marked increase from the fifth and sixth to the seventh and eighth grades. Similar though less conspicuous changes occur in inhibition, but here population Z shows only a slight increase from the lower to the higher classes.

The retarded children, chiefly because they are older, tend to be more persistent than the rest; and when all populations are combined, the accelerated, chiefly because they are younger, appear to be less persistent than the normal. Very little relation is found between inhibition and grade status.

Teachers' academic marks reflect persistent conduct as measured by our tests only slightly. (But in one school [Z] scholastic marks correlate .431 with inhibition on three tests.) The *r*'s between deportment and inhibition indicate that in estimating pupils' good conduct the power of inhibition is taken into consideration. (These *r*'s are .258, .314, and .306.)

2. Sex Differences. Extreme differences between the two sexes occur on the inhibition tests, the girls proving far better able to resist the distractions and temptations presented by all the situations used. (In terms of their significance, the differences between boys and girls on the separate tests are: Pictures, 2.4; Puzzles, 5.5; Safes, 4.8; Stories, 6.4; total, 5.5.) On two out of five persistence tests the girls worked enough harder than the boys to bring their total persistence score to a point slightly above that of the boys. This superiority of the girls is probably due largely to differences in interest — the girls care less than boys for mechanical puzzles and are more eager to get good marks. Hence they find it easier to let the puzzles and safes alone and are better able to resist the distractions of the Picture test.

3. Physical and Emotional Condition. No relation between physical condition, as measured by our athletic ability tests, and persistence is found, while the *r*'s between physical condition and inhibition are low (.173, .132, and .076). The same lack of relation is noted in the case of physical condition as measured by status of nutrition, eyes, and teeth. Children who are extremely strong for their age, however, are also well inhibited, and in population Y the extremely inhibited are significantly stronger than the extremely uninhibited.

In the case of emotional condition, the fact of nail biting seems to indicate something of importance, as those who make a practice of doing this are far less persistent than those who leave their nails alone. The latter are also somewhat better inhibited on our tests. The emotional stability score on the Woodworth-Mathews test shows an association with both persistence and inhibition similar to that found between intelligence and these behaviors. But a large group of children reporting a very large number of nervous

symptoms also scored very high on the persistence tests. When it is recalled that the most persistent group is the foreign and less favored, which also shows bigger neurotic scores, this peculiarity can be understood.

4. The General Environment. As just noted, the less favored community and the children of foreign born in all communities are more persistent than the rest and better inhibited on the Pictures and Stories tests. Y showed its superiority on the Persistence for Self and for Class tests and was better able to let the safes alone than the rest. X excelled on the Story Resistance test and the Magic Square test and broke even with Y on the Puzzle Inhibition test and with Z on the Story Inhibition test. Similar differences appear between the children of native and foreign-born parents, but the differences among the religious groups are very small.

Social-cultural level is slightly associated with higher persistence and inhibition scores (the Burdick test correlating with the former .146, .248, and .148, and with the latter .097, .132, and .322). These cut across the differences in community and national background just reported and move in the opposite direction. Within any one community, social background is somewhat associated with self-control, but the association is not strong enough to counterbalance the effect of other factors.

The net result of this section is to emphasize the fact that conduct is a function of specific situations and experiences.

5. The Home. Brothers and sisters resemble one another in self-control about as in service and to a lesser extent than in deception or intelligence. (The r's for separate tests average .212 for persistence and .265 for inhibition. The r's for combinations of tests reach .458 for persistence and .344 for inhibition.) These resemblances cannot be accounted for by reference to the influence of such home background features as we have measured and may represent, in part at least, inherited biological factors.

6. The School. Such influence as schools exert seems to be through the medium of the classroom morale, which operates independently of the length of time a child is associated with the school or community. The growth of morale seems to be specifically

related to experience in the school and community, improving in some situations and retrograding in others. Disruptive influences, such as a teacher's loss of discipline, seem to have a generally unfavorable effect, but it is probable that constructive influences have a less general favorable effect.

7. Out-of-School Experience. Length of Sunday-school attendance is somewhat more definitely related to persistence than is length of day-school attendance but is quite unrelated to inhibition. (The r's, with age constant, are .169, .241, and .087 for persistence. The observed r's are $-.004$, .079, and $-.005$ for inhibition.) In all populations, those not enrolled have a higher persistence average than those enrolled, and the same is true of the inhibition scores in the case of population Y. Regularity of attendance shows no constant relation to self-control.

Club membership is not associated with superior self-control scores. Those who belong to the three groups called systems Y_1, Y_2, and Y_3 are just like other children of their own age and sex in these types of conduct. Similarly, attendance at summer camps shows no effect on the ability to score on the tests. But in the case of motion picture attendance there was observed a tendency for those attending relatively less than their schoolmates to exhibit greater effort and greater power of self-restraint.

8. Miscellaneous Concomitants. Resistance to the type of suggestion included in the Otis test correlates slightly with persistence (.283, .231, and .118), but not with inhibtion (.060, .079, and .192). Social suggestion apparently operates to make pupils in some rooms give up working on difficult tasks in groups rather than as individuals — that is, when the act of quitting is openly noticeable.

A substantial difference ($2\frac{1}{2}$ times its standard error) is found between the persistence scores of children whose skin shows no signs of " old dirt " and those of children who are less well groomed, and a smaller difference (just equal to its standard error) is found in inhibition scores. In each case the difference favors the cleaner children.

Sociability, in the sense of being named by others as a friend, goes

with moderate persistence scores and with moderate or high inhibition scores. The influence of friendship, in the sense of association, is measured by the correlation between friends within and without the classroom. Resemblance in persistence is about the same as in service tendencies, but the resemblance in inhibition is considerably less. (For total scores, the r's in the case of persistence are .417, .206, and .293, and in the case of inhibition, .115, −.001, and .223.) These resemblances are of pairs of friends both of whom are in each case within the same classroom. The corresponding r's for friends who are not in the same classroom are usually lower.

Sustained-speed ability is involved in certain of our tests which use speed material. (The r's between this ability and the Persistence for Self test are .193, .185, and .178, but they are about zero in the case of the Persistence for Class test.) The Picture Inhibition test correlates fairly with sustained-speed ability (.293, .377, and .655), and the indications are that the ability to make high scores on speed tests of the sort used is partly a function of the ability to resist distractions. No correlation between speed and the other inhibition tests is found.

Being with the majority sex in a classroom may have some bearing on conduct. In one population the correlation between the percentage of boys in a class and the ratio of boys' scores to girls' scores is .85 for persistence and .30 for inhibition. The effect is similar to that reported for service and dishonesty.

Summary. It is hazardous to attempt any summary of these miscellaneous associations. What we do give must be recognized as quite inadequate, since almost any statement made about this complex of factors should have numerous qualifications attached to it. Depending on the reader's good sense to supply the cautions which the abbreviated form of presentation must omit, we offer the following:

Individual Differences. Inhibition and persistence represent certainly a balance of interests and probably also specific abilities or habits of work. In our persistence tests we presented no alternatives to each task save more tasks of like character. In our inhibition

tests we arranged a double situation, and the resulting behavior indicated to which part of the situation the pupil was most responsive or best able to respond. Thus in the Safes test we placed the attractive object within reach, but we so modified this situation as to make any manipulation of the safe unfair to others. If the safe was handled, it was at the expense of fairness to classmates and also, of course, in opposition to the direct request of the examiner.

That persistence is chiefly a matter of interest seems to be indicated by the relation between scores on the persistence tests and age, intelligence, and sex. Other factors may also be involved, however, for the scores increase with age even with children of the same intelligence. Furthermore, nervous stability and power to resist social suggestion are associated with high persistence score. The resemblance of friends and siblings might be interpreted as a reflection of common interests, for certainly such facts as culture and economic status show only the slightest association with persistence.

Inhibition, on the other hand, which involves a more obvious conflict of interests, does not present such a clear picture. The outstanding instance of individual difference is sex. This would seem to indicate greater sensitiveness on the part of girls to those subtle aspects of the situation calling for inhibition and greater sensitiveness on the part of boys to the grosser and more obvious aspects of the situation, such as manipulation possibilities in the puzzles and safes. The inhibition scores are not greatly influenced by age or intelligence or emotional condition or suggestibility or home background, but friends and siblings tend to resemble one another, suggesting that the ability to respond to a somewhat abstract appeal may be a feature which has some biological significance and which enters into the choice of friends.

Group Differences. The largest group differences on the various self-control tests are between national groups and communities, the less favored proving the more persistent in two out of three kinds of test and better inhibited in three out of four tests. Those who do not attend Sunday school seem to belong to those portions

of the population who do best on the persistence tests, and the same is usually true of those who attend motion pictures less than their schoolmates. This does not necessarily imply that Sunday schools and motion pictures make children less persistent, for in both cases the children who attend less may do so for reasons associated with the causes of their superiority in persistence. This seems quite plausible in view of the fact that it is the children of the foreign born who score highest on two of these tests — and they probably have less money for pictures and are less interested in Sunday schools. So far as our records show, there is a slight tendency (somewhat stronger than in the case of day schools) for those longer affiliated with Sunday schools to be more persistent.

Membership in clubs for boys and girls makes no difference in self-control on our tests, but being in one classroom rather than another is of considerable importance, even though day schooling in terms of years in one school is always without significance.

There is no indication at all that on any one test or any combination of tests children will fall into two groups — the controlled and the uncontrolled. On the contrary, the majority are at or near an average of self-control, and there are a few who are very poorly controlled and a few who are extremely well controlled.

IMPLICATIONS

The implications one derives from studies of the sort here presented depend largely on the general trend of his educational philosophy. Some will infer one thing and some another. Doubtless hostile critics of certain conventional procedures in school and church will find comfort in some of our findings. By the same token, the supporters of the educational theory which leans heavily on the joint experience of a group with the realities to be learned will discover evidence which favors their point of view. We have endeavored not to take sides in any educational controversy in the accumulation and organization of our data. It would be very easy to destroy the value of any such investigation as this by the exercise of partisanship in the type of data sought or in its presentation.

Presumably we are not free from bias, but at points where we have found ourselves pleased with the results that were appearing we have been especially cautious to make sure that the facts were free from error and that the evidence for opposing views was pictured even more completely than for views with which we were in sympathy on other grounds.

The conclusions as to the facts have already been presented. What do these facts mean for education?

We find that the implications of our *Studies in Deceit* and our *Studies in Service* (Book One of this volume) are borne out in this investigation of persistence and inhibition. There is little evidence that effectively organized moral education has been taking place. There is abundant evidence, however, that children have been acquiring habits which are important for character. The picture of American childhood which our report presents is far from discouraging. We shall summarize it later, after we have presented the next two steps in our study in Volume III. But what they are at present learning of self-control, as also of service and honesty, is largely a matter of accident. Peculiarities of home, church, school, Sunday school, teacher, club leader, and everything else that deliberately attempts to influence the child work upon him by divers means and with divers results. Anarchy in the leadership of moral education is not likely to produce order in the character of a child. At all events, such leadership as we have in typical American communities has not resulted in organized conduct.

This seems at first to be a counsel of despair. But the explanation we have offered of how children actually learn honesty, service, and self-control, which is in harmony with the general trend of the psychology and philosophy of education, gives us confidence in the possibility of moral education. The absence of it is due to the operation of the very causes which make it possible. Or, to put the matter another way, the wise educator will work with the laws of learning as he finds them operating, not against them. Briefly, then, the most important implications we see in our study are as follows:

1. What is to be learned must be experienced.

2. What is to be experienced must be represented in the situations to which children are exposed.

3. If what is to be learned is some form of conduct or mode of adjustment, then the situations to which children are exposed must be opportunities to pursue interests which lead to the conduct to be learned.

4. This conduct must be carried on in relation to the particular situations to which it is the preferred mode of response.

5. A common and potent factor in such situations is the established practice and code of the group, which by coloring the situation may either hinder or assist the acquisition of desirable responses on the part of its members.

6. If standards and ideals, whether already in the possession of the group or not, are to function as controlling factors, they must become a part of the situation to which the child responds and assist in the achievement of satisfactory modes of ad ustment to those aspects of the situation which are independent of these standards and ideals — they must be tools rather than objects of esthetic appreciation.

7. The achievement of specific standards, attitudes, and modes of conduct does not imply their integration. Integration is itself a specific achievement.

APPENDICES

		PAGE
A.	Details of Technique Used in Scaling Helpful Behavior	455
B.	(1) Directions for the Efficiency Coöperation and Free Choice Tests; (2) Pupil Data Sheet	459
C.	Supplementary Tables: (1) Referred to in Book One; (2) Referred to in Book Two	466
D.	Computation of CAVI Sigma Scores	527
E.	Directions for Administering Self-Control Tests as Given in Populations X, Y, and Z	529
F.	Results of Self-Control Testing in Population C	537
G.	(1) Standardization of Picture Inhibition Test Material; (2) Stability of the Derived Picture Inhibition Score	545
H.	Directions for Inhibition Parties as Given in Population C	548

APPENDIX A

DETAILS OF TECHNIQUE USED IN SCALING HELPFUL BEHAVIOR

Cheating attitudes seem to be quite specific. The scales work for any one type of situation but do not cut across the situations.

So far as we have attempted to scale helpful behavior, the attitudes seem to be somewhat more general. At least two different types of helpful behavior will fall into a common scale, *viz.*, giving up money and giving up ice cream.

The general method of scale building is discussed in Chapter XII, Book Two, *Studies in Deceit.* This difference between the operation of the two attitudes, however, should be noted. The cheating attitude operates presumably as a highly complex pattern involving various ulterior motivations and various conflicting codes. Furthermore, it operates against social resistance, offered by means of disapproval and supervision.

Helpfulness, on the other hand, would seem to be relatively simple in its pattern, less dependent on ulterior motives, and more directly related to its object as its end. Also it has the approval and coöperation of the community. It is like cheating in that it does something to someone, just as honesty is like refusing to help in that it refrains from doing something to someone. The thing that is scaled is not the absence of the behavior but the behavior, and so cheating rather than honesty, and helpfulness rather than selfishness, are handled in the scores and scales.

TABLE A

RESULTS OF ICE CREAM AND MONEY GIFT TESTS — POPULATION D

	Number Giving All Cash (3.6%)	Number Voting All, 1st or 2d Vote (2.5%)	Number Giving Part Cash (12.5%)	Number Voting Part, 1st or 2d Vote (21%)	Number Giving Ice Cream, 1st Vote (28%)	Number Giving Ice Cream, 2d Vote (27%)	Number Giving Ice Cream, 3d Vote (1.8%)	Per Cent of Spaces Filled (1)	(2)
	7	7	*	*	7	*	*		
	1	1	*	*	0	1	*	. . 98	96
		3	3	*	3	*	*		
		2	0	*	2	*	*	. . 93	87
			19	19	19	*	*		
			6	6	0	6	*		
			2	2	0	0	0	. . 90	88
			1	0	0	1	*		
				29	29	*	*		
				12	0	12	*		
				1	0	0	1	. . 85	81
				5	0	0	0		
					63	*	*		
						60	*	. . 100	100
							4	94	91
Totals .	8	13	31	74	123	80	5	215	
Per cent of steps filled .	100	100	94	98	82	91	42	88%	

Experimental study of the percentages of pupils in population D who gave something at the various stages of each of the two occasions described in Chapter I of this volume resulted in the combinations and arrangement of Table A, which is equivalent to Table CLXXIII of the chapter in *Studies in Deceit* just referred to.

The stars in the table represent spaces logically filled, as those who gave all their money of course gave part, and those voting ice cream on the lowest appeal naturally are included as the appeal was increased. The column called (1) under "Per Cent of Spaces Filled" shows the extent to which the scale operates if these logically filled spaces are regarded as realities. Thus for the first two rows there are $7 \times 7 + 1 \times 7$, or 56, possible vacancies, all but one of which, or 98%, are filled. The column labeled (2) shows the percentage of spaces filled when only those spaces are included which represent real decisions. Thus for the first two rows there are $7 + 7 + 7 + 1 + 1 + 1$, or 24, possible vacancies, all but one of which, or 96%, are filled. Thus 91% of all possible spaces are occupied. The figures at the bottom of the table show the extent to which each of the several steps is a true step. Thus for the first step, 5 out of a total of 4 + 5 (for the first 0) + 1 + 2 (for the second 0), or 12, pupils are scaled, which equals 42%.

The following processes correspond to those described fully in *Studies in Deceit*. Table B gives the cumulative percentages reaching the several steps of the scale and the equivalent SD's.

TABLE B

SD Values for Steps of Scale

Helpful Act	%	SD	Scale Value
All cash	3.6	1.80	48
All votes	5.8	1.57	46
Part cash	18.3	0.90	39
Part votes	39.4	0.27	33
Ice cream 1	67.6	− 0.46	25
Ice cream 2	94.6	− 1.61	14
Ice cream 3	96.4	− 1.80	12

These are translated into score values by reference to the following scale, which calls − 3 SD zero and each tenth SD one:

− 3.0 SD	− 2.5	− 2.0	− 1.5	− 1.0	− 0.5	0.0	+1.5	+ 1.0	+ 1.5	+2.0	+ 2.5	+ 3.0
0	5	10	15	20	25	30	35	40	45	50	55	60

Showing the present levels in italics, the scale runs as follows:

0 1 2 3 . . . 11 *12* 13 *14* 15 . . . 24 *25* 26 . . . 32 *33* 34 . . . 38 *39* 40 . . . 45 *46* 47 *48* 49 . . . 60

The particular acts of helpfulness represented in the money test and the ice cream test thus fall at the points italicized on a scale reaching from an absolute zero of just not any helpfulness to a maximum score of 60. It is not probable that those who refused to give up their cream on a third appeal were at this point of absolute zero. Indeed the percentages would seem to indicate that they are not far below those who waited for the third appeal before giving up their ice cream. Assuming that their distance below the third appeal is about the same as the distance of the third below the second, we have 10 as the score value of the zeros. The scores for each step now read:

10 12 14 25 33 39 46 48

For our purposes those who did not give on the third appeal may be called zero, which makes the scale read:

0 2 4 15 23 29 36 38

This divided by two, for convenience, is:

0 1 2 7 11 14 18 19

Putting these numbers back into their proper places in the series, we have:

0 1 2 3 4 5 6 *7* 8 9 10 *11* 12 13 *14* 15 16 17 *18 19*

In order to take care of those who skipped one or more intervals of the scale, we evaluate the steps by reference to their distance from one another on a theoretical scale in which every step is occupied by a test. Table C gives these theoretical levels and the corresponding score value for each of the actual steps.

The scaled helpfulness score now becomes the sum of the score values of the helpful acts exhibited by the subject.

The reliability of such a scale can be determined only by repeated test-

ing of the same children. Its self-consistency is measured by the extent to which its steps and spaces are filled, as shown in Table A.

TABLE C

SCORE VALUES OF HELPFULNESS SCALE

HELPFUL ACT	LEVEL	STEP	SCORE
Giving nothing	0		0
Ice cream 3	1	1	1
Ice cream 2	2	1	1
	3	1	
	4	1	
	5	1	
	6	1	
Ice cream 1	7	1	5
	8	1	
	9	1	
	10	1	
Part votes	11	1	4
	12	1	
	13	1	
Part cash	14	1	3
	15	1	
	16	1	
	17	1	
All votes	18	1	4
All cash	19	1	1
Highest possible score			19

Helpfulness score = sum of score values of the things done

APPENDIX B

1. DIRECTIONS FOR THE EFFICIENCY COÖPERATION AND FREE CHOICE TESTS

I

Pass out the tests, serial number down, and say: "Do not touch until I give you directions."

After all have their copies, say: "I will explain this work after we have had two minutes' practice in addition. Do not write your name on this first sheet. It will not count on your score. It is for practice only.

When I say, 'Go,' write the word 'Practice' at the top and begin adding, and work until I tell you to stop. Ready? Go."

After 60 seconds say: "Stop. When I say, 'Go,' tear off this first sheet and continue adding on the other side. (Demonstrate. Be sure that tearing is in unison.) Ready? Go."

After 60 seconds say: "Stop. Turn this sheet upside down and lay it at one side. You will need it later.

"Now we are going to have two speed contests. One contest is to find out who is the fastest worker in this class, the second fastest, third fastest, and so on to the very slowest. The other contest is to find out which *class* in this school is fastest, second fastest, third fastest, and so on. Each of you has 20 sheets of simple additions. Some of them are for one contest and some for the other. To make it more interesting, you will change on every sheet. On one sheet you will work for a prize for yourself, and on the other sheet you will work for a prize for your class. You will write your own name on those papers on which you work for yourself. You will write the words 'Class ____' (fill in) on those sheets on which you work for your class.

"The score of each one will be the number of additions done on all the papers on which you work for yourself and on which you write your name. The scores for the class will be the number of additions done on all the papers which all of you do for the class. All the papers marked 'Class ____' will be kept separate. A prize will be given those of you making good scores on your own papers, and every one of you has a chance to win a prize. A class prize will be given the class which is the fastest, that is, the class which gets the most examples done. Your class has a good chance to win this class prize.

"Now turn to the first sheet. You will note that the first addition is $2 + 3$. On this sheet you will work for yourself. When I say, 'Go,' write your name at the top and begin to add. The faster you work, the higher will be your own score. Ready? Go."

After exactly 60 seconds say: "Stop. Pencils up. When I say, 'Go,' tear off this first sheet, turn it over, and continue adding for yourself on the other side. Ready? Go."

After 60 seconds say: "Stop. Put this sheet aside, upside down, and turn to the next sheet. On this sheet you will work for your class. When I say, 'Go,' write the words 'Class ____' at the top and begin adding. The faster you work on this sheet, the higher will be your class score. But everything you do counts. Ready? Go."

After 60 seconds say: "Stop. When I say, 'Go,' tear off this sheet, turn it over, and continue adding for your class. Ready? Go."

After 60 seconds say: "Stop. Put this sheet to one side, upside down, and turn to the next sheet. On this sheet you will work for yourself again. When I say, 'Go,' write your name at the top and begin adding. Ready? Go."

After 60 seconds say: "Stop. When I say, 'Go,' tear off this sheet and continue adding for yourself on the other side. Ready? Go."

After 60 seconds say: "Stop. Put this sheet to one side, upside down, and turn to the next sheet. On this one you will work for your class prize again. When I say, 'Go,' write the words 'Class ____' at the top and begin adding. Ready? Go."

After 60 seconds say: "Stop. When I say, 'Go,' tear off this sheet and continue adding for your class prize on the other side. Ready? Go."

After 60 seconds say: "Stop. Now put your two class papers together. Be sure your name is not on them. Will the last person in each row please come forward and collect these class papers? Keep the papers with your own name on, and the practice sheet. The score for the class is the total number of examples done on these sheets and on the others which you will do for the class later."

Take class sheets, shuffle them conspicuously, place them to one side, and say: "Now pass the two sheets on which you have placed your own name. Be sure your name is on each paper." Place these in a separate place and say, "Each one's score is the total number of examples he does on his own paper."

II

Say: "Now we are ready to continue these two contests. On the next 8 sheets you will again work on one sheet for yourself and on the other for your class, as before. Just as before, each one's score will be the total number of additions done on all the papers on which you work for yourself and on which you write your name. All the papers marked 'Class ____' will be put in one bunch. The score of this class will be the total number of additions done on all these class papers. Don't forget that a prize will be given to those of you who make good scores on your own papers, and a class prize will be given to the fastest class, that is, the class which gets the most examples done on the papers marked 'Class ____.'

"Now turn to the first sheet. On this sheet you will work for yourself. When I say, 'Go,' write your name at the top and start to add. The faster you work, the higher will be your score. Ready? Go."

After 60 seconds say: "Stop. When I say, 'Go,' tear off this sheet and continue adding for yourself on the other side. Ready? Go."

After 60 seconds say: "Stop. Put this sheet to one side, upside down, and turn to the next sheet. On this sheet you will work for your class. When I say, 'Go,' write the words 'Class ____' at the top and start to work. Ready? Go."

After 60 seconds say: "Stop. When I say, 'Go,' tear off this sheet, turn it over, and continue adding for your class on the other side. Ready? Go." Same directions on next six sheets, alternately.

III

Say: "There are seven more sheets. You may work on as many of these as you wish for yourself and as many as you wish for your class. You may even do all the papers for yourself or all for the class. Whatever you do for yourself will count on your own score, and whatever you do for your class will add to the class score.

"At the end of the contest I will collect all the papers at one time. Look at the next sheet. Remember, if you write your name on this sheet, it will count for yourself, for your own score. If you write the words 'Class ____' on it, it will count for your class. When I say, 'Go,' write either your name or the words 'Class ____' on the sheet and begin to add. Ready? Go."

After 60 seconds say: "Stop. Put this sheet to one side. We shall not use the other side of it. Turn to the next sheet. When I say, 'Go,' write either the words 'Class ____' or your own name at the top, according as you wish the examples to count for your class or yourself, and begin to work. Ready? Go."

After 60 seconds, as usual.

Give directions like the preceding for the next four sheets, alternating the order in which class and self are referred to.

Then say: "Now we come to the very last sheet of the contest. The score on this sheet will count for more than any sheet so far. The score on this sheet will count for both you and for your class; so put your name at the top and write the words 'Class ____' also. Remember, this is your last chance to make a high score in speed. Work as fast as ever you can. Ready? Go."

After 60 seconds say: "Stop. When I say, 'Go,' turn over the sheet and continue working as rapidly as possible. This is your very last minute of work. Ready? Go."

After 60 seconds say: "Stop. Lay this last sheet aside. Now find your practice sheet. You may add this sheet either to your score or to the score of your class. If you want it for yourself, put your name on it. If you want it for the class, write the words 'Class ____' on the sheet."

Allow time; then have papers collected.

2. PUPIL DATA SHEET

Name.....Date.....
School.....Grade.....

1. On the lines below write the names and grades of the playmates or companions you like best. Give both first and last names. If you can't spell a name, guess at the spelling.
   1. Name.....Grade.....
   2. Name.....Grade.....
   3. Name.....Grade.....
   4. Name.....Grade.....
2. What do you usually do on Saturdays? Put down everything, both play and work.
   .....
   .....
   .....
   .....
   .....
   .....
3. What do you usually do on Sundays? Put down everything.
   .....
   .....
   .....
   .....
   .....
   .....
4. About how many times a week do you go to the movies?.....
5. What kinds of moving pictures do you like best?.....
   .....
6. What moving picture you have seen lately have you thought about a good deal?.....
   .....
7. About how many times a month do you go to dances, parties, etc.?.....
8. What sort of books do you like best to read?.....
9. What book or books have you read this past month just for the fun of it?
   .....
   .....

10. What magazine do you read regularly?.....
.....
11. Which part of this magazine do you like best?.....
.....
12. Since you started going to school what things have you most enjoyed playing with?.....
.....
.....
.....
13. What are you most afraid of?.....
.....
14. Since you first started to school you have had several teachers. Which ones have you liked best?
    1. ..... 3. .....
    2. ..... 4. .....
15. Which ones have you disliked or not got along with?
    1. ..... 3. .....
    2. ..... 4. .....
16. Are you ever punished for doing something you know you ought not to have done?.....
    About how often?.....By whom?.....
    What kinds of punishment do you the most good?.....
    .....
    What kinds of punishment do *not* do you any good?.....
    .....
    .....
17. Do you earn any money?.....How much a week?.....
    How do you earn it?.....
18. How much do you save a week?.....
19. Does anyone give you an allowance each week?.....
    .....
    If so, who?.....How much?.....
20. How many brothers have you?.....
    1. Name.....Age.....School.....Grade.....
    2. Name.....Age.....School.....Grade.....
    3. Name.....Age.....School.....Grade.....
21. How many sisters have you?.....
    1. Name.....Age.....School.....Grade.....
    2. Name.....Age.....School.....Grade.....
    3. Name.....Age.....School.....Grade.....
22. What country was your father born in?.....
23. What country was your mother born in?.....
24. What church does your father go to?.....

25. What church does your mother go to?.....
26. What clubs and societies do you now belong to, such as Rangers, Bluebirds, Junior Scouts, Boy Scouts, Girl Scouts, Camp Fire Girls, Girl Reserves, Pioneers, Christian Endeavor, and the like?
    1. Name of club.....Name of leader.....
    Address of leader.....
    When did you join?.....Are you now an active member?.....
    If not now a member, when did you drop out?.....
    .....
    Where are meetings held?.....
    At what time during the week are meetings held?
    1st meeting: Day.....Hour.....
    2d meeting: Day.....Hour.....
    What merit badges, beads, tokens, or other recognitions have you won?
    ....
    ....
    ....
    What rank have you achieved (*e.g.*, tenderfoot, second class, woodgatherer, etc.)?
    .....
    2. Name of club.....Name of leader.....
    Address of leader.....
    When did you join?.....Are you now an active member?.....
    If not now a member, when did you drop out?.....
    .....
    Where are meetings held?.....
    At what time during the week are meetings held?.....
    1st meeting: Day.....Hour.....
    2d meeting: Day.....Hour.....
    What merit badges, beads, tokens, or other recognitions have you won?
    ....
    ....
    ....
    What rank have you achieved (*e.g.*, tenderfoot, second class, woodgatherer, etc.)?
    .....
27. What summer camp for boys or girls did you attend last summer or the summer before?
    .....
    Address of camp.....
    Name of director.....
    How many summers have you been going?.....
    Camp honors won.....
    .....

Offices held.....
What special interest did you follow?.....
.....
What part of camp life did you like best?.....
.....
What part of camp life did you like least?.....
.....
Did anything happen to you that you very greatly disliked?.....
What was it?.....
.....
.....
.....
.....
.....

APPENDIX C

SUPPLEMENTARY TABLES

1. REFERRED TO IN BOOK ONE

TABLE

I. Median Total Service Score for Each Age Level — Populations X, Y, and Z

II. Total Service Scores by Grade — Populations X, Y, and Z

III. Efficiency Coöperation Scores by Grade — Populations X, Y, and Z

IV. Money Vote Scores by Grade — Populations X, Y, and Z

V. Average Number of Examples Given to Class by Each Grade on the Free Choice Test — Populations X, Y, and Z

VI. Median Service Scores of Accelerated, Normal, and Retarded Children, by Grade and Population

VII*A*. Median Scores on the Free Choice Test of Accelerated, Normal, and Retarded Children, by Grade

VII*B*. Median Scores on the Free Choice Test of Accelerated, Normal, and Retarded Children, by Population

VIII. Median Scores on the Efficiency Coöperation Test of Accelerated, Normal, and Retarded Children, by Grade

IX. Total Service Scores of Pupils Accelerated, Normal, and Retarded, with Reference to Age-Group Status — Population Y

X. Scores on the Free Choice Test of Pupils Accelerated, Normal, and Retarded, with Reference to Age-Group Status — Population Y

TABLE

XI. Scores on the Efficiency Coöperation Test of Pupils Accelerated, Normal, and Retarded, with Reference to Age-Group Status — Population Y

XII. Scores on the Money Vote Test of Pupils Accelerated, Normal, and Retarded, with Reference to Age-Group Status — Population Y

XIII. Kits Test Scores of Pupils Accelerated, Normal, and Retarded, with Reference to Age-Group Status — Population Y

XIV. Envelopes Test Scores of Pupils Accelerated, Normal, and Retarded, with Reference to Age-Group Status — Population Y

XV. Total Service Reputation Scores of Pupils Accelerated, Normal, and Retarded, with Reference to Age-Group Status — Population Y

XVI. Distribution of Service Medians by Years for Four Classes of Populations X and Z

XVII. Median Scores on Free Choice, Efficiency Coöperation, and Money Vote Tests for Bright, Average, and Dull Pupils

XVIII. Per Cent of Bright, Average, and Dull Pupils Coöperating on Kits and Envelopes Tests

XIX. Sex Differences in the Free Choice Test — Populations X, Y, and Z

XX. Sex Differences in the Efficiency Coöperation Test — Populations X, Y, and Z

XXI. Sex Differences during the Efficiency Coöperation Test — Populations X, Y, and Z

XXII. Sex Differences in the Average Number of Examples Done for Self and for Class in Each Two-Minute Period of the Efficiency Coöperation Test — Populations X, Y, and Z Combined

XXIII. Sex Differences on the Money Vote Test — Populations X, Y, and Z

XXIV. Sex Differences on the Kits Test — Populations X, Y, and Z

XXV. Sex Differences on the Envelopes Test — Populations X, Y, and Z

XXVI. Sex Differences in Total Service Score — Populations X, Y, and Z

XXVII. Sex Differences in "Guess Who" and Check List Scores Relating to Helpfulness — Populations X, Y, and Z

XXVIII. Sex Differences in Total Service Reputation Score — Populations, X, Y, and Z

XXIX. Sex Differences in Total Service Score, by Grade — Populations X, Y, and Z

XXX. Population Differences in Chronological Age (Months)

TABLE
XXXI. Population Differences in Intelligence (CAVI Sigma)
XXXII. Population Differences in Emotional Stability
XXXIII. Population Differences in Physical Ability
XXXIV. Population Differences in Physical Condition
XXXV. Population Differences in Teachers' Marks
XXXVI. Population Differences in Deportment
XXXVII. Population Differences in "Guess Who" and Check List Scores
XXXVIII. Population Differences in Service Reputation Score
XXXIX. Population Differences in Total Service Score
XL. Population Differences in the Free Choice Test
XLI. Population Differences in the Efficiency Coöperation Test
XLII. Population Differences in the Money Vote Test
XLIII. Population Differences in the Kits Test
XLIV. Population Differences in the Envelopes Test
XLV. Occupational Differences in Total Service Score
XLVI. Occupational Differences on the "Self *versus* Class" Tests
XLVII. Occupational Differences on the Charity Tests
XLVIII. Culture Score (Burdick) in Relation to Service Score
XLIX. Service Scores of Children Whose Parents Were Born in the United States and of Children Whose Parents Were Foreign Born
L. National Differences in Total Service Score
LI. National Differences on Separate Service Tests
LII. Service Scores of Children of Different Religious Groups — All Populations Combined
LIII. Religious Differences in Burdick and Sims Scores
LIV. Comparison of Actual Service Scores of Each Religious Group with Scores Expected on Basis of Cultural Differences
LV. Religious Differences on Separate Service Tests — Population X
LVI. Classroom Service Means and Sigmas
LVII. Test of Presence of Factors Unequally Influencing Service Tendencies of Classrooms
LVIII. Percentage Distributions of Service Reputation Scores of Selfish and Unselfish Boys and Girls, by Population
LIX. Percentage Distributions of Physical Condition Scores of Selfish and Unselfish Pupils of Population Y
LX. Percentage Distributions of Burdick Culture Scores of Selfish and Unselfish Pupils, by Population
LXI. Percentage Distributions of Movie Attendance of Selfish and Unselfish Pupils, by Population
LXII. Percentage Distributions of Intelligence (CAVI Sigma) Scores of Selfish and Unselfish Pupils, by Population

TABLE

LXIII. Percentage Distributions of Emotional Stability Scores of Selfish and Unselfish Pupils, by Population

LXIV. Percentage Distributions of Ages of Selfish and Unselfish Pupils, by Population

2. REFERRED TO IN BOOK TWO

LXV. Median Persistence Score for Each Age Level — Populations X, Y, and Z

LXVI. Median Inhibition Score for Each Age Level — Populations X, Y, and Z

LXVII. Total Persistence Scores by Grade — Populations X, Y, and Z

LXVIII. Story Resistance Scores by Grade — Populations X, Y, and Z

LXIX. Ring and Cross Scores by Grade — Populations X, Y, and Z

LXX. Magic Square Scores by Grade — Populations X, Y, and Z

LXXI. Persistence for Self Scores by Grade — Populations X, Y, and Z

LXXII. Persistence for Class Scores by Grade — Populations X, Y, and Z

LXXIII. Total Inhibition Scores by Grade — Populations X, Y, and Z

LXXIV. Total Inhibition Scores without the Picture Test, by Grade — Populations X, Y, and Z

LXXV. Persistence Medians for Age and Grade

LXXVI. Persistence for Self Medians of Accelerated, Normal, and Retarded Pupils — Population Z

LXXVII. Inhibition Medians (Three Tests Only) by Age and Grade

LXXVIII. Sex Differences on the Story Resistance Test and the Ring and Cross Test

LXXIX. Sex Differences on the Magic Square Test

LXXX. Sex Differences on the Persistence for Self and for Class Tests

LXXXI. Sex Differences in Total Persistence Score and in Reputation for Persistence

LXXXII. Sex Differences on the Picture Inhibition Test

LXXXIII. Sex Differences on the Puzzle Manipulation Test

LXXXIV. Sex Differences on the Safe Manipulation Test

LXXXV. Sex Differences on the Story Inhibition Test and Total Inhibition Score

LXXXVI. Sex Differences in Reputation for Inhibition

LXXXVII. Population Differences in Total Persistence Score

LXXXVIII. Population Differences on the Story Resistance Test

LXXXIX. Population Differences on the Ring and Cross Test

XC. Population Differences on the Magic Square Test

XCI. Population Differences on the Persistence for Self Test

TABLE
XCII. Population Differences on the Persistence for Class Test
XCIII. Population Differences in Total Inhibition Score
XCIV. Population Differences in Puzzles + Safes + Stories
XCV. Population Differences on the Picture Inhibition Test
XCVI. Population Differences on the Puzzle Manipulation Test
XCVII. Population Differences on the Safe Manipulation Test
XCVIII. Population Differences on the Story Inhibition Test
XCIX. National Differences in Total Persistence Score
C. National Differences on Separate Persistence Tests
CI. National Differences in Total Inhibition Score
CII. National Differences on Separate Inhibition Tests
CIII. Classroom Persistence Scores
CIV. Classroom Inhibition Scores
CV. Per Cent of Overlap of the Nonpersistent on the Persistent Group — Populations X, Y, and Z
CVI. Per Cent of Overlap of the Noninhibited on the Inhibited Group — Populations X, Y, and Z
CVII. Probability of Occurrence of Differences between Nonpersistent and Persistent Groups — Populations X, Y, and Z
CVIII. Probability of Occurrence of Differences between Noninhibited and Inhibited Groups — Populations X, Y, and Z

SUPPLEMENTARY TABLES

1. REFERRED TO IN BOOK ONE

TABLE I

MEDIAN TOTAL SERVICE SCORE FOR EACH AGE LEVEL — POPULATIONS X, Y, AND Z

AGE	POPULATION X		POPULATION Y		POPULATION Z	
	N	Median	N	Median	N	Median
16	1		4	130.0		
15	3	107.5	13	113.0	9	98.3
14	6	100.0	24	115.7	20	105.6
13	21	112.1	45	117.9	63	106.5
12	48	111.1	64	117.2	57	110.8
11	76	108.8	51	114.8	52	114.4
10	43	113.2	45	114.3	25	115.9
9	18	116.0	25	118.5		
Total		111		116		109

TABLE II

TOTAL SERVICE SCORES BY GRADE — POPULATIONS X, Y, and Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	36	112.0	113.05	8.90	96	115.8	116.51	10.47	46	115.8	113.05	8.90
6	66	113.2	113.40	11.40	67	113.0	113.40	9.96	46	110.0	113.40	11.40
7	75	107.6	108.16	8.46	74	116.3	117.36	10.30	133	108.2	108.16	8.46
8	42	113.3	113.10	12.45	39	122.5	123.52	14.28	82	106.1	113.10	12.45
5 and 6	102	112.6	113.28	10.58	163	114.4	115.23	10.38	92	113.4	113.28	10.58
7 and 8	117	109.1	109.94	10.35	113	118.1	119.49	12.18	215	107.1	109.94	10.35
Total	219	110.6	112.03	10.42	276	115.8	117.03	11.44	307	108.9	108.04	13.12

GRADE	POPULATION XYZ			
	N	Median	Mean	SD
5	178	114.9	115.80	10.12
6	179	112.4	112.80	10.89
7	282	109.9	119.79	12.48
8	163	110.9	111.51	15.12
5 and 6	357	113.7	114.42	10.58
7 and 8	445	110.2	110.44	13.59
Total	802	112.9	112.10	11.42

TABLE III

EFFICIENCY COÖPERATION SCORES BY GRADE* — POPULATIONS X, Y, AND Z

GRADE	POPULATION X			POPULATION Y			POPULATION Z		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
5	37	8.3		97	10.0		46	8.8	
6	65	6.7		68	8.7		47	6.8	
7	77	8.0		73	8.7		133	5.0	
8	39	6.8		40	8.8		87	3.7	
5 and 6	102	7.3	3.8	165	9.5	2.8	93	7.8	4.5
7 and 8	116	7.6	4.9	113	8.8	3.9	220	4.5	9.0

* The raw score on this test is the average difference between the examples

done per minute for self and the examples done per minute for class. In this table the raw score is subtracted from 10 for convenience of presentation and to correspond to Figures 6, 7, and 8. A mean of 6.8 represents a mean difference of 3.2 examples between self and class, in favor of self.

TABLE IV

MONEY VOTE SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X			POPULATION Y			POPULATION Z		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
5	33	6.4		93	4.8		46	5.9	
6	68	6.5		64	4.4		46	5.2	
7	76	4.8		72	5.3		127	4.8	
8	44	5.7		37	5.5		88	5.0	
5 and 6	101	6.5	2.7	157	4.6	1.9	92	5.5	2.6
7 and 8	120	5.1	2.5	109	5.4	2.2	215	4.9	2.2

TABLE V

AVERAGE NUMBER OF EXAMPLES GIVEN TO CLASS BY EACH GRADE ON THE FREE CHOICE TEST — POPULATIONS X, Y, AND Z*

GRADE	POPULATION X		POPULATION Y		POPULATION Z		TOTAL	
	N	Mean	N	Mean	N	Mean	N	Mean
5	37	1.85	97	2.31	46	3.48	180	2.51
6	66	2.06	68	2.51	45	2.26	179	2.28
7	76	1.72	74	3.65	133	2.83	283	2.75
8	42	3.14	40	2.85	87	2.78	169	2.88
5 and 6	103	1.99	165	2.40	91	2.86	359	2.40
7 and 8	118	2.23	114	3.37	220	2.81	452	2.80

* There were seven exercises in all, and each pupil decided for himself what proportion he would give to his class for a class prize and what proportion to himself for an individual prize. The figures are the mean number of exercises given to class out of these seven.

TABLE VI

MEDIAN SERVICE SCORES OF ACCELERATED, NORMAL, AND RETARDED CHILDREN, BY GRADE AND POPULATION

GRADES IN X	ACCELERATED		NORMAL		RETARDED	
	N	Median	N	Median	N	Median
5	15	111.9	23	114.5	11	109.2
6	23	112.5	44	111.2	7	113.8
7	16	106.7	16	104.4	6	107.5
8	28	115.0	19	111.9	6	97.5
GRADES IN Y						
5	27	118.8	25	116.1	35	112.5
6	13	110.5	16	115.8	32	114.5
7	13	118.1	23	114.6	30	116.2
8	8	122.5	14	127.5	17	120.8
GRADES IN Z						
5	—	—	29	114.2	24	116.9
6	5	112.5	18	113.3	32	104.4
7	23	108.8	40	110.0	77	107.0
8	20	108.6	35	107.5	26	102.2
GRADES IN XYZ						
5	42	116.5	77	114.9	70	113.5
6	41	111.7	78	112.7	71	111.6
7	52	110.0	79	110.9	113	109.3
8	56	113.6	68	111.4	49	104.3

TABLE VIIA

MEDIAN SCORES ON THE FREE CHOICE TEST OF ACCELERATED, NORMAL, AND RETARDED CHILDREN, BY GRADE

GRADE	ACCELERATED		NORMAL		RETARDED	
	N	Median	N	Median	N	Median
5	39	2.2	78	2.3	70	2.3
6	42	1.6	77	2.6	69	1.4
7	52	2.7	78	3.1	112	2.6
8	58	2.8	70	2.5	50	1.4
XYZ	191	2.3	303	2.7	301	1.9

TABLE VIIB

MEDIAN SCORES ON THE FREE CHOICE TEST OF ACCELERATED, NORMAL, AND RETARDED CHILDREN, BY POPULATION

POPULATION	ACCELERATED		NORMAL		RETARDED	
	N	Median	N	Median	N	Median
X	79	1.9	105	3.0	28	2.4
Y	61	2.3	88	2.0	124	1.3
Z	51	2.8	110	2.8	149	2.2
XYZ	191	2.3	303	2.7	301	1.9

TABLE VIII

MEDIAN SCORES ON THE EFFICIENCY COÖPERATION TEST OF ACCELERATED, NORMAL, AND RETARDED CHILDREN, BY GRADE

GRADE	ACCELERATED		NORMAL		RETARDED	
	N	Median	N	Median	N	Median
5	43	8.5	74	9.3	71	8.9
6	39	8.3	79	8.4	70	7.0
7	52	8.3	75	8.8	104	6.7
8	58	8.1	56	7.6	49	7.1
XYZ	192	8.3	284	8.5	294	7.5

TABLE IX

TOTAL SERVICE SCORES OF PUPILS ACCELERATED, NORMAL, AND RETARDED, WITH REFERENCE TO AGE-GROUP STATUS — POPULATION Y

	N	MEDIAN	MEAN	SD
Accelerated boys	11	118.3	115.9	10.0
Normal boys	107	115.7	116.9	11.0
Retarded boys	22	113.8	114.1	9.5
Accelerated girls	13	119.0	118.8	8.4
Normal girls	107	117.1	118.1	12.6
Retarded girls	16	115.7	115.6	11.4
Accelerated boys and girls	24	118.8	117.5	9.2
Normal boys and girls	214	116.4	117.5	11.9
Retarded boys and girls	38	114.7	114.7	10.4
Retarded and accelerated boys	33	115.0	114.7	9.7
Retarded and accelerated girls	29	117.1	117.1	10.3
Retarded and accelerated boys and girls	62	116.1	115.8	10.0

TABLE X

SCORES ON THE FREE CHOICE TEST OF PUPILS ACCELERATED, NORMAL, AND RETARDED, WITH REFERENCE TO AGE-GROUP STATUS — POPULATION Y

	N	MEDIAN	% COÖPERATING
Accelerated boys	11	3.5	91
Normal boys	108	2.6	70
Retarded boys	22	2.3	76
Accelerated girls	13	1.8	69
Normal girls	108	2.0	72
Retarded girls	17	3.2	83
Accelerated boys and girls	24	3.0	79
Normal boys and girls	216	2.3	71
Retarded boys and girls	39	2.8	80
Retarded and accelerated boys	33	2.9	82
Retarded and accelerated girls	30	2.8	77
Retarded and accelerated boys and girls	63	2.9	79

TABLE XI

SCORES ON THE EFFICIENCY COÖPERATION TEST OF PUPILS ACCELERATED, NORMAL, AND RETARDED, WITH REFERENCE TO AGE-GROUP STATUS — POPULATION Y

	N	MEDIAN
Accelerated boys	11	9.6
Normal boys	107	9.8
Retarded boys	23	9.8
Accelerated girls	13	9.2
Normal girls	107	8.8
Retarded girls	17	6.9
Accelerated boys and girls	24	9.4
Normal boys and girls	214	9.3
Retarded boys and girls	40	8.6
Retarded and accelerated boys	34	9.7
Retarded and accelerated girls	30	7.9
Retarded and accelerated boys and girls	64	8.9

TABLE XII

SCORES ON THE MONEY VOTE TEST OF PUPILS ACCELERATED, NORMAL, AND RETARDED, WITH REFERENCE TO AGE-GROUP STATUS — POPULATION Y

	N	MEDIAN
Accelerated boys	11	4.6
Normal boys	109	4.6
Retarded boys	20	5.0
Accelerated girls	12	5.5
Normal girls	101	5.1
Retarded girls	13	5.5
Accelerated boys and girls	23	4.9
Normal boys and girls	210	4.8
Retarded boys and girls	33	5.3
Retarded and accelerated boys	31	4.8
Retarded and accelerated girls	25	5.5
Retarded and accelerated boys and girls	56	5.1

TABLE XIII

KITS TEST SCORES OF PUPILS ACCELERATED, NORMAL, AND RETARDED, WITH REFERENCE TO AGE-GROUP STATUS — POPULATION Y

	N	% COÖPERATING
Accelerated boys	9	56
Normal boys	116	67
Retarded boys	24	54
Accelerated girls	13	85
Normal girls	106	61
Retarded girls	17	65
Accelerated boys and girls	22	73
Normal boys and girls	222	64
Retarded boys and girls	41	59
Retarded and accelerated boys	33	55
Retarded and accelerated girls	30	73
Retarded and accelerated boys and girls	63	64

TABLE XIV

ENVELOPES TEST SCORES OF PUPILS ACCELERATED, NORMAL, AND RETARDED, WITH REFERENCE TO AGE-GROUP STATUS — POPULATION Y

	N	% COÖPERATING
Accelerated boys	11	55
Normal boys	116	48
Retarded boys	23	35
Accelerated girls	13	85
Normal girls	113	73
Retarded girls	14	64
Accelerated boys and girls	24	71
Normal boys and girls	229	60
Retarded boys and girls	37	46
Retarded and accelerated boys	34	41
Retarded and accelerated girls	27	74
Retarded and accelerated boys and girls	61	62

TABLE XV

TOTAL SERVICE REPUTATION SCORES OF PUPILS ACCELERATED, NORMAL, AND RETARDED, WITH REFERENCE TO AGE-GROUP STATUS — POPULATION Y

	N	MEAN	SD
Accelerated boys	10	.90	9.5
Normal boys	110	.65	11.4
Retarded boys	22	−.14	15.4
Accelerated girls	13	10.96	9.5
Normal girls	106	7.02	4.5
Retarded girls	14	5.57	6.1
Accelerated boys and girls	23	6.58	10.7
Normal boys and girls	216	3.79	9.2
Retarded boys and girls	36	2.08	13.0
Retarded and accelerated boys	32	.19	13.9
Retarded and accelerated girls	27	8.17	8.3
Retarded and accelerated boys and girls	59	3.84	12.3

TABLE XVI

DISTRIBUTION OF SERVICE MEDIANS BY YEARS FOR FOUR CLASSES OF POPULATIONS X AND Z

TOTAL SERVICE SCORES

AGE	Z, GRADE 5		X, GRADE 5		Z, GRADE 6		X, GRADE 8	
	N	Median	N	Median	N	Median	N	Median
8			1	102.5				
9			13	112.5	1	92.5		
10	21	114.2	24	114.2	3	127.5		
11	13	116.2	8	108.3	15	115.0	4	108.3
12	7	117.5	1	102.5	17	106.2	22	115.4
13	3	117.5			6	105.0	19	110.1
14					1	117.5	5	92.5
15					3	97.5		
16							1	102.5

FREE CHOICE SCORES

8			1	2.5				
9	1	1.5	14	2.0	1	2.0		
10	21	3.3	25	1.2	3	3.8		
11	13	3.3	8	1.0	16	2.0	4	3.0
12	7	4.8	1	0.5	17	1.5	22	2.7
13	3	3.5			7	1.0	18	2.1
14					1	1.5	5	1.8
15					2	2.0		
16							1	0.5

EFFICIENCY COÖPERATION SCORES

8			1	3.0				
9	1	9.0	14	9.1	1	1.0		
10	21	8.8	25	9.3	3	10.5		
11	13	9.5	8	8.3	16	8.0	4	9.0
12	7	10.8	1	9.0	17	6.8	22	7.5
13	3	8.5			7	7.0	18	9.2
14					1	17.0	5	− 1.0
15					2	1.0		
16							1	7.0

TABLE XVII

MEDIAN SCORES ON FREE CHOICE, EFFICIENCY COÖPERATION, AND MONEY VOTE TESTS FOR BRIGHT, AVERAGE, AND DULL PUPILS

FREE CHOICE TEST

POPULATION	BRIGHT		AVERAGE		DULL	
	N	Median	N	Median	N	Median
X	58	1.8	121	1.9	27	1.6
Y	38	3.0	197	2.5	36	2.3
Z	35	2.7	180	2.3	72	2.1
XYZ	131	3.3	498	2.5	135	2.3

EFFICIENCY COÖPERATION TEST

X	59	8.0	122	7.9	26	7.5
Y	38	10.7	198	9.2	36	10.5
Z	34	8.1	180	7.3	73	6.3
XYZ	131	8.8	500	8.3	135	7.6

MONEY VOTE TEST

X	59	6.0	120	6.2	27	5.5
Y	38	4.9	176	4.7	31	5.4
Z	30	4.9	160	5.2	70	5.4
XYZ	127	5.3	456	5.2	128	5.4

TABLE XVIII

PER CENT OF BRIGHT, AVERAGE, AND DULL PUPILS COÖPERATING ON KITS AND ENVELOPES TESTS

KITS TEST

POPULATION	BRIGHT		AVERAGE		DULL	
	N	%	N	%	N	%
X	61	67	125	64	28	36
Y	38	68	194	63	40	60
Z	17	18	128	48	72	39
XYZ	116	60	447	59	140	44

ENVELOPES TEST

X	60	25	131	36	29	28
Y	40	70	204	61	38	50
Z	34	59	184	38	75	28
XYZ	134	47	519	47	142	34

TABLE XIX

SEX DIFFERENCES IN THE FREE CHOICE TEST — POPULATIONS X, Y, AND Z

POPULATION		N	MEDIAN	MEAN	SD	% GIVING ALL	% SHARING	% GIVING NOTHING
X	Boys	114	1.3	1.8	1.5	1	56	43
	Girls	107	2.2	2.5	1.7	1	77	22
Y	Boys	141	2.6	2.9	2.2	10	62	28
	Girls	138	2.1	2.6	2.0	3	70	27
Z	Boys	162	2.5	2.8	2.2	6	64	30
	Girls	149	2.7	2.8	1.7	4	79	17
XYZ	Boys	417		2.6	2.1	12	57	31
	Girls	394		2.7	1.8	6	73	21

TABLE XX

SEX DIFFERENCES IN THE EFFICIENCY COÖPERATION TEST* — POPULATIONS X, Y, AND Z

POPULATION		N	MEDIAN	MEAN	SD
X	Boys	117	8.4	7.2	4.8
	Girls	101	8.1	6.5	4.0
Y	Boys	141	9.9	9.5	2.5
	Girls	137	9.5	8.4	3.8
Z	Boys	163	7.5	6.0	7.7
	Girls	150	7.2	5.6	8.4
XYZ	Boys	421	8.8	7.8	5.9
	Girls	388	8.5	7.0	6.2

* The score is the number of examples done per minute when working for self minus the number per minute when working for class. This figure is subtracted from 10 in this table, so that greater coöperation appears as a higher score. The SD is the sigma of the raw difference between self and class.

TABLE XXI

SEX DIFFERENCES DURING THE EFFICIENCY COÖPERATION TEST* — POPULATIONS X, Y, AND Z

POPULATION	SUCCESSIVE TWO-MINUTE PERIODS						
	1	2	3	4	5	6	
X	1.1	0.9	1.2	2.4	5.2	5.9	Boys
Y	− 1.4	− 0.2	0.5	0.3	3.0	1.4	
Z	4.9	3.4	2.1	0.9	5.2	4.0	
XYZ	1.3	1.2	1.2	1.8	4.4	4.1	
X	1.8	− 0.3	1.5	3.1	5.6	7.9	Girls
Y	− 0.3	0.6	1.7	1.7	2.9	3.3	
Z	4.6	1.3	3.0	2.5	6.1	9.6	
XYZ	1.6	0.6	2.1	2.3	4.6	5.5	

* The score is the number of examples done per minute when working for self *in excess* of the number done when working for class. Thus in population

X, the boys did on the average 1.1 more examples for self than for class per minute during the first two minutes of the test and 5.9 more for self than for class per minute during the last two minutes. In Y, however, both boys and girls did more for class than for self during the first period.

TABLE XXII

SEX DIFFERENCES IN THE AVERAGE NUMBER OF EXAMPLES DONE FOR SELF AND FOR CLASS IN EACH TWO-MINUTE PERIOD OF THE EFFICIENCY COÖPERATION TEST — POPULATIONS X, Y, AND Z COMBINED

	SUCCESSIVE TWO-MINUTE PERIODS						
	1	2	3	4	5	6	
Examples for self . . .	42.6	42.2	43.3	43.4	43.8	43.1	Boys
Examples for class . .	41.3	41.0	42.1	41.6	39.4	39.0	
Examples for self . . .	49.1	48.7	50.0	49.8	49.7	50.5	Girls
Examples for class . .	47.5	48.1	47.9	47.5	45.1	44.0	

TABLE XXIII

SEX DIFFERENCES ON THE MONEY VOTE TEST — POPULATIONS X, Y, AND Z

POPULATION	N	MEDIAN	MEAN	SD
X { Boys . .	119	5.7	5.5	2.8
X { Girls . .	102	6.2	6.0	2.5
Y { Boys . .	140	4.7	4.6	2.0
Y { Girls . .	126	5.2	5.3	2.1
Z { Boys . .	160	4.9	4.9	2.2
Z { Girls . .	147	5.4	5.2	2.4
XYZ { Boys . .	419		5.0	2.4
XYZ { Girls . .	375		5.5	2.3

TABLE XXIV

SEX DIFFERENCES ON THE KITS TEST — POPULATIONS X, Y, AND Z

POPULATION		N	% GIVING	MEAN OF GIVERS	SD OF GIVERS
X	Boys	122	61	5.4	4.2
	Girls	107	64	6.1	4.1
Y	Boys	149	64	7.4	6.5
	Girls	136	64	6.2	5.2
Z	Boys	128	41	4.5	2.2
	Girls	114	39	4.0	1.6
XYZ	Boys	399	56	6.1	5.2
	Girls	357	56	5.7	4.4

TABLE XXV

SEX DIFFERENCES ON THE ENVELOPES TEST — POPULATIONS X, Y, AND Z

POPULATION		N	% GIVING	MEAN OF GIVERS	SD OF GIVERS
X	Boys	122	25	43.4	18.7
	Girls	109	36	43.5	20.1
Y	Boys	150	47	35.8	18.6
	Girls	140	73	45.5	19.8
Z	Boys	176	26	36.7	19.0
	Girls	151	44	42.5	20.5
XYZ	Boys	448	33	37.7	19.0
	Girls	400	52	44.1	20.1

TABLE XXVI

SEX DIFFERENCES IN TOTAL SERVICE SCORE — POPULATIONS X, Y, AND Z

POPULATION		N	MEDIAN	MEAN	SD
X	Boys	112	109	110.4	11.1
	Girls	107	112	112.6	10.1
Y	Boys	140	115	116.3	10.8
	Girls	136	116	117.6	11.9
Z	Boys	161	108	107.6	12.2
	Girls	146	110	109.1	14.0
XYZ	Boys	413		111.3	11.9
	Girls	389		113.0	12.9

TABLE XXVII

SEX DIFFERENCES IN "GUESS WHO" AND CHECK LIST SCORES RELATING TO HELPFULNESS — POPULATIONS X, Y, AND Z

POPULATION		"GUESS WHO"					CHECK LIST	
		N	Median	Mean	SD	% +	N	% +
X	Boys	117	1.0	1.20	6.62	60	124	84
	Girls	105	2.5	2.80	5.18	77	112	95
Y	Boys	144	.9	.02	9.60	54	147	57
	Girls	137	3.4	4.64	8.84	88	144	83
Z	Boys	167	.8	− .64	8.04	57	165	62
	Girls	132	3.5	4.28	5.46	87	155	76
XYZ	Boys	428		.18	8.16	57	436	66
	Girls	374		4.00	5.50	85	411	84

TABLE XXVIII

SEX DIFFERENCES IN TOTAL SERVICE REPUTATION SCORE — POPULATIONS X, Y, AND Z

POPULATION		N	MEDIAN	MEAN	SD*	% +
X	Boys	116	4.1	3.95	9.85	76
	Girls	107	5.7	5.71	7.15	87
Y	Boys	142	2.1	.55	13.15	58
	Girls	133	6.8	7.25	5.90	93
Z	Boys	161	3.1	1.35	10.90	65
	Girls	145	7.4	7.35	7.75	88
XYZ	Boys	419		1.80	11.35	66
	Girls	385		7.00	7.00	89

* The scores range from − 40 to + 40.

TABLE XXIX

SEX DIFFERENCES IN TOTAL SERVICE SCORE, BY GRADE — POPULATIONS X, Y, AND Z

BOYS

POPULATION	GRADE 5		GRADE 6		GRADE 7		GRADE 8	
	N	Mean	N	Mean	N	Mean	N	Mean
X	22	112.8	34	112.5	37	108.4	19	108.1
Y	54	105.0	36	114.5	31	117.0	19	121.0
Z	26	115.4	22	107.9	69	106.4	44	104.8
XYZ	102	109.3	92	112.2	137	109.3	82	109.3

GIRLS

X	14	113.6	32	114.0	38	107.9	23	117.3
Y	42	118.0	31	111.9	43	117.5	20	125.5
Z	20	118.5	24	114.4	64	106.6	38	105.1
XYZ	76	117.3	87	113.4	145	110.2	81	113.6

TABLE XXX

Population Differences in Chronological Age (Months)

By School

Population	N	Median	Mean	SD
X	243	138.5	139.8	15.4
Y	299	144.4	146.0	21.2
Z	340	153.4	152.8	16.3

$Z - Y = 4.5\,\sigma$
$Y - X = 4.0\,\sigma$

By Grade and School

Grade	Population X			Population Y			Population Z			Order of Means
	Number of Classes	Mean	SD	Number of Classes	Mean	SD	Number of Classes	Mean	SD	
5	1	122.4	8.5	3	132.5	17.2	1	132.7	11.3	(ZY)X
6	2	131.9	9.3	2	146.0	14.9	1	145.7	15.6	(YZ)X
7	2	145.0	14.2	2	155.0	14.0	3	159.5	12.0	Z Y X
8	1	155.8	12.9	2	170.1	15.2	2	162.0	13.0	Y Z X
Total	6	139.8	15.4	9	146.0	21.2	7	152.8	16.3	

TABLE XXXI

Population Differences in Intelligence (CAVI Sigma)

Population	N	Median	Mean	SD
X	222	50.65	50.80	9.75
Y	286	47.50	47.44	7.92
Z	302	45.50	45.43	8.25

$X - Y = 4.2\,\sigma$
$Y - Z = 3.0\,\sigma$

TABLE XXXII

Population Differences in Emotional Stability

Woodworth-Mathews Personal Data Sheet

Population	N	Median	Mean	SD
X	217	12.80	15.47	9.07
Y	270	16.81	17.78	10.18
Z	288	21.06	22.04	10.48

Lower scores indicate greater stability.

$X - Y = 2.65\,\sigma$
$Y - Z = 4.9\,\sigma$

Neurotic Index Based on CEI Pupil Data Sheet

Population	N	Median	Mean	SD
X	217	2.0	2.44	1.94
Y	283	2.1	3.01	2.92
Z	298	2.0	2.90	2.89

Lower scores indicate greater stability.

$X - Y = 2.62\,\sigma$

TABLE XXXIII

Population Differences in Physical Ability

By School

Population	N	Median	Mean	SD
X	221	107.3	109.64	18.11
Y	276	102.0	103.08	24.36
Z	313	114.3	111.88	19.98

$X - Y = 3.4\,\sigma$
$Z - Y = 4.8\,\sigma$

By Grade and School

Grade	Population X		Population Y		Population Z		Order of Means
	Mean	SD	Mean	SD	Mean	SD	
5	99	16	93	17	100	16	(ZX)Y
6	102	14	100	19	110	15	Z X Y
7	110	15	109	26	118	19	Z(XY)
8	126	20	121	29	127	20	(ZX)Y

TABLE XXXIV

Population Differences in Physical Condition

Population	N	Median	Mean	SD
X	219	101.42	99.98	15.66
Y	274	100.20	98.30	16.65
Z	310	107.43	107.00	16.80

$Z - Y = 5.2\,\sigma$

TABLE XXXV

Population Differences in Teachers' Marks

Population	N	Mean	SD
X	210	82.9	8.7
Y	264	79.6	7.6
Z	282	75.3	10.7

TABLE XXXVI

Population Differences in Deportment

Population	N	Median	Mean	SD
X	226	9.5	9.02	1.99
Y	281	8.7	8.05	3.11
Z	306	8.5	8.07	2.29

TABLE XXXVII

Population Differences in "Guess Who" and Check List Scores

"Guess Who"					Check List
Population	N	Median	Mean	SD	% above Zero
P	102	2.8	3.20	14.60	
X	222	2.0	1.96	6.00	89
Y	281	2.7	2.28	8.20	70
Z	299	2.2	1.52	7.46	68

TABLE XXXVIII

Population Differences in Service Reputation Score

Population	N	Median	Mean *	SD	% Positive
X	223	5.25	5.06	8.50	81
Y	275	5.13	3.80	10.80	75
Z	306	5.63	4.20	10.00	76

* The scale runs from − 40 to + 40.

TABLE XXXIX

Population Differences in Total Service Score

Population	N	Median	Mean	SD
X	219	111	111.50	10.4
Y	276	116	116.95	11.4
Z	307	109	108.35	13.1

TABLE XL

Population Differences in the Free Choice Test

Population	N	% Giving at Least One Exercise to Class	Median of All	Mean of Givers Only
X	221	68	1.8	2.9
Y	279	73	2.7	3.6
Z	311	76	2.6	3.5

TABLE XLI

Population Differences in the Efficiency Coöperation Test

Population	N	Median	Mean	SD
X	221	8.4	7.44	4.4
Y	279	9.8	9.20	3.3
Z	311	7.3	5.46	8.0

TABLE XLII

Population Differences in the Money Vote Test

Population	N	Median	Mean	SD
P	244	6.9	6.3	2.4
X	221	6.0	5.7	2.6
Y	266	4.9	4.9	2.1
Z	307	5.2	5.1	2.3
XYZ	794		5.2	2.3

TABLE XLIII

Population Differences in the Kits Test

Population	N	% Contributing	Median of All	Mean Amount of Givers	SD of Givers
P	164	83	7.0	10.4*	11.2
X	229	62	3.2	5.8	4.7
Y	285	64	2.7	6.8	6.0
Z	242	40	0.8	4.3	2.0
XYZ	756	56		5.9	5.0

* The slight differences in the values assigned to the articles in the kits in population P as compared with population XYZ would not affect the scores enough to alter materially the difference between the mean of P and the means of X, Y, and Z.

TABLE XLIV

Population Differences in the Envelopes Test

Population	N	% Coöperating	Median of All	Mean of Coöperators	SD of Coöperators
P	164	34.0			
X	231	30.7	00	43.4	20.2
Y	290	59.3	33.2	41.6	19.9
Z	321	34.3	00	40.3	20.0
XYZ	842	42.1		41.5	20.0

TABLE XLV

OCCUPATIONAL DIFFERENCES IN TOTAL SERVICE SCORE

OCCUPATIONAL LEVEL	POPULATION X		POPULATION Y		POPULATION Z	
	N	Median	N	Median	N	Median
I	55	112.5	6	120.0	0	—
II	60	111.0	20	120.0	14	111.2
III	61	110.1	109	115.4	47	107.5
IV	13	109.2	96	116.3	156	110.2
V	5	102.5	18	113.6	20	104.2
All		111		116		109

TABLE XLVI

OCCUPATIONAL DIFFERENCES ON THE "SELF *VERSUS* CLASS" TESTS

POPULATION X

OCCUPATIONAL LEVEL	FREE CHOICE			EFFICIENCY COÖPERATION	
	N	% Coöperating	Median	N	Median
I	57	72	1.7	57	7.9
II	60	70	1.9	60	8.1
III	63	65	1.9	64	8.4
IV	13	69	1.8	13	7.5
V	5	40	0.7	5	5.5

POPULATION Z

II	16	88	3.2	16	8.6
III	44	82	2.4	46	7.9
IV	164	76	2.8	160	7.5
V	20	75	2.6	19	5.6

TABLE XLVII

Occupational Differences on the Charity Tests

Occupational Level	Money Vote		Kits		Envelopes			
	N	Median	N	% Coöperating	N	% Coöperating	Number of Givers	Median of Givers
Population X								
I	60	6.2	61	61	62	40	25	50.0
II	61	6.3	62	68	63	30	19	42.3
III	63	5.9	63	59	65	28	18	33.5
IV	13	5.8	14	86	14	7	1	—
V	8	6.0	8	37½	8	12½	1	—
Population Z								
II	15	4.5	8	75	19	47	9	43.5
III	52	5.2	44	34	47	32	15	46.0
IV	159	5.3	126	38	171	39	67	42.9
V	17	5.4	20	40	22	27	6	14.3

TABLE XLVIII

Culture Score (Burdick) in Relation to Service Score

Culture Scores	Population X		Population Y		Population Z		XYZ	
	N	Service Median	N	Service Median	N	Service Median	N	Service Median
276–325	23	116.1	3	123.8	0		26	117.2
216–275	81	111.4	27	118.8	15	109.2	123	111.1
156–215	60	110.4	66	117.5	95	112.8	221	113.5
96–155	23	107.5	127	115.1	125	108.5	275	112.3
36–95*	2	102.5	21	112.1	39	105.2	62	107.3
All	189	111	244	116	274	109	707	112

* One case in X and one in Y fall in the next lower interval.

TABLE XLIX

Service Scores of Children Whose Parents Were Born in the United States and of Children Whose Parents Were Foreign Born

Test	U. S.			U. S. and Foreign			Foreign		
	N	Median	% Giving	N	Median	% Giving	N	Median	% Giving
Total service score	430	113.8		71	112.0		194	108.6	
Free Choice	431	2.2	72	71	2.5	77	199	2.4	72
Efficiency	431	9.9		71	8.3		199	6.9	
Kits and Envelopes	384	6.4	74	73	3.2	60	199	2.2	52
Vote + Choice + Efficiency	382	107.0		72	106.3		192	105.3	

TABLE L

National Differences in Total Service Score

Nationality of Parents	Population X		Population Y		Population Z		Population XYZ	
	N	Median	N	Median	N	Median	N	Median
Slavic* and U. S.	11	109.4			1	127.5	12	110.0
Italian and U. S.	5	107.5	1	112.5	7	103.8	13	106.3
German and U. S.	1	112.5	6	117.5	10	112.5	17	114.2
Scandinavian and U. S.	1	127.5	3	107.5			4	117.5
English and U. S.	1	107.5	9	117.5	1	122.5	11	117.5
Irish and U. S.	2	97.5	3	101.2	7	113.8	12	105.0
Irish	2	100.0	2	115.0	22	108.3	26	106.7
English			1	102.5	3	113.8	4	112.5
Scandinavian	1	117.5	1	122.5	1	107.5	3	117.5
French and Belgian	1	107.5	3	122.5	1	107.5	5	108.8
German and Austrian	1	117.5	8	110.0	9	111.2	18	111.0
Italian	7	110.8	6	112.5	59	103.2	72	105.5
Slavic	5	108.3			40	109.1	45	109.5
Jewish	8	113.3	2	122.5				
Different			1	112.5	3	117.5		
No record	38	110.3	23	114.2	48	110.6	109	110.8
Total		111		116		109		112

* In population X the Slavic group is Jewish in all mixed cases.

TABLE LI

National Differences on Separate Service Tests

Free Choice

Nationality of Parents	Population X			Population Y			Population Z			Population XYZ		
	N	Median	% Giving	N	Median	% Giving	N	Median	% Giving	N	Median	% Giving
U. S.	136	1.7	69	209	2.5	74	86	2.6	75	431	2.2	72
U. S. and foreign	21	1.8	57	22	3.0	86	28	2.8	86	71	2.5	77
Foreign	27	2.3	70	23	1.9	65	149	2.4	74	199	2.4	72
U. S. and Italian	5	1.5	60	1	0.0	0	7	1.5	57	13	1.3	54
U. S. and Irish	2	1.5	50	3	1.2	67	7	4.5	100	12	2.0	83
U. S. and Scandinavian	1	3.5	100	3	2.5	67				4	3.0	50
U. S. and English	1	3.5	100	9	3.8	100	1	2.5	100	11	3.6	100
Italian	7	2.2	72	6	1.3	67	60	2.0	68	73	1.9	68
Irish	2	0.5	0	2	4.5	100	22	2.4	73	26	2.3	69
Scandinavian	1	5.5	100	1	3.5	100	1	4.5	100	3	4.5	100
English				1	0.0	0	4	5.0	100	5	3.5	80

Money Vote, Free Choice, and Efficiency*

	N	Median	N	Median	N	Median	N	Median
U. S. and Italian	5	105.0	1	99.0	7	102.5	13	102.5
U. S. and Irish	2	95.0	3	99.5	7	114.3	12	103.3
U. S. and Scandinavian	1	109.0	3	107.0			4	108.0
U. S. and English	1	103.0	10	114.0	1	107.0	12	113.0
Italian	7	107.5	6	104.0	61	100.5	74	103.3
Irish	2	100.0	2	109.0	22	106.0	26	105.0
Scandinavian	1	109.0	1	109.0	1	105.0	3	109.0
English			1	103.0	4	113.0	5	111.0

Kits and Envelopes*

	N	Median	% Giving	N	Median	% Giving	N	Median	% Giving	N	Median	% Giving
U. S. and Italian	5	1.7	40	1	15.0	100	7	1.7	43	13	1.9	46
U. S. and Irish	2	4.0	100	3	1.5	33	7	1.4	29	12	1.7	42
U. S. and Scandinavian	1	17.0	100	2	2.0	50				3	3.0	67
U. S. and English	1	5.0	100	9	7.0	78	1	15.0	100	11	7.0	82
Italian	7	3.0	71	5	1.7	67	60	1.9	48	72	2.0	51
Irish	2	2.0	50	2	3.0	50	21	1.7	43	25	1.8	44
Scandinavian	1	9.0	100	1	13.0	100	1	3.0	100	3	9.0	100
English				1	1.0	0	3	10.5	67	4	5.0	50

* Cf. Table XLIX for U. S. and foreign totals.

TABLE LII

Service Scores of Children of Different Religious Groups — All Populations Combined

Religion	N	Median	Mean	SD
Protestant	288	115.3	115.97	11.28
Catholic	238	108.3	107.00	12.60
Jewish	48	111.0	112.70	8.60
Mixed	44	114.3	114.88	13.30
No affiliation	38	111.7	113.02	14.09
No record	176	111.2	111.39	10.30

Protestant — Catholic, 8.5 σ
Protestant — Jewish, 2.3 σ
Jewish — Catholic, 3.8 σ

TABLE LIII

Religious Differences n Buridick and Sims Scores

Burdick

Religion	Population X			Population Y			Population Z			Total		
	N	Mean	SD	N	Median	Mean	N	Median	Mean	N	Mean	SD
Protestant	109	227	49	179	148	157	27	159	159	315	181.4	57.5
Jewish	43	239	44	4	132	132				47	229.6	51.9
Catholic	61	192	53	19	130	142	180	145	144	260	154.8	51.2

Sims

Protestant	111	26.4	4.9	183	16.5	17.2				294	20.7	7.2
Jewish	40	25.6	4.2	4	21.0	20.5				44	25.1	4.4
Catholic	64	20.2	6.6	24	16.5	16.1				88	19.1	6.5

Jewish cultural — Protestant cultural, 5.9 σ
Protestant cultural — Catholic cultural, 5.9 σ
Jewish Sims — Protestant Sims, 5.6 σ
Protestant Sims — Catholic Sims, 2.0 σ

TABLE LIV

COMPARISON OF ACTUAL SERVICE SCORES OF EACH RELIGIOUS GROUP WITH SCORES EXPECTED ON BASIS OF CULTURAL DIFFERENCES

POPULATION X

RELIGION	BURDICK MEAN	EXPECTED SERVICE MEDIAN	ACTUAL SERVICE MEDIAN	DIFFERENCE BETWEEN ACTUAL AND EXPECTED MEDIANS
Protestant . .	227	111.5	112.6	+ 1.1
Jewish	239	111.1	109.8	− 1.3
Catholic . . .	192	111.7	107.9	− 3.8

POPULATION Y

Protestant . .	157	115.3	117.2	+ 1.9
Jewish	132	115.4	122.5	+ 7.1
Catholic . . .	142	115.0	112.9	− 2.1

POPULATION Z

Protestant . .	159	110.5	110.6	+ 0.1
Catholic . . .	144	108.9	107.8	− 1.1

POPULATION XYZ

Protestant . .	181	113.3	115.3	+ 2.0
Jewish	230	111.9	111.0	− 0.9
Catholic . . .	155	111.9	108.3	− 3.6

TABLE LV

RELIGIOUS DIFFERENCES ON SEPARATE SERVICE TESTS — POPULATION X

FREE CHOICE

RELIGION	N	% GIVING	MEAN	SD
Protestant . .	118	69	2.2	1.9
Jewish	42	71	2.0	1.4
Catholic . . .	61	64	1.7	1.2

EFFICIENCY COÖPERATION

	N	MEAN	SD	
Protestant . .	115	7.8	5.5	
Jewish	42	7.0	6.9	
Catholic . . .	61	6.9	5.5	

MONEY VOTE

	N	MEAN	SD	
Protestant . .	123	7.7	2.1	
Jewish	39	6.2	2.6	
Catholic . . .	59	5.0	2.7	

	KITS		ENVELOPES	
	N	% GIVING	N	% GIVING
Protestant . .	126	67	124	39
Jewish	41	62	43	28
Catholic . . .	63	52	64	17

14977

TABLE LVI

Classroom Service Means and Sigmas

Population X

Class	Grade	N	Mean	SD	Median		$\frac{\sigma}{\sqrt{N}}$
					Boys	Girls	
H	5	36	113.1	8.9	111.6	112.5	1.48
I	6	32	112.2	10.8	108.2	117.5	1.90
J	6	34	114.6	11.9	116.5	112.9	2.03
K	7	35	110.0	8.0	107.5	110.8	1.35
L	7	40	106.6	8.6	106.7	106.0	1.35
M	8	42	113.1	12.5	109.4	117.5	1.92
			$\sigma = 3.3$				Average 1.67
							$\frac{3.3}{1.67} = 1.83$

Population Y

N	5	29	114.4	7.0	114.0	114.5	1.30
O	5	33	119.8	13.6	113.0	123.1	2.37
P	5	34	115.2	8.7	115.0	117.5	1.48
Q	6	29	115.3	9.8	114.0	113.5	1.81
R	6	38	112.0	10.1	113.3	111.7	1.63
S	7	34	116.8	10.9	113.2	122.5	1.86
T	7	40	117.9	10.0	118.8	115.0	1.57
U	8	14	119.0	15.6	126.3	113.8	4.16
V	8	25	126.1	12.8	121.7	131.3	2.55
			$\sigma = 3.8$				Average 2.08
							$\frac{3.8}{2.08} = 1.98$

TABLE LVI (*Continued*)

POPULATION Z

CLASS	GRADE	N	MEAN	SD	MEDIAN		$\frac{\sigma}{\sqrt{N}}$
					Boys	Girls	
A	5	46	117	9.0	114.3	117.5	1.33
B	6	46	111	11.2	107.1	113.0	1.64
C	7	45	112	12.2	109.2	115.0	1.81
D	7	46	104	14.1	106.3	105.5	2.07
E	7	42	103	12.7	105.9	105.5	1.95
F	8	41	110	9.4	108.9	111.3	1.46
G	8	41	96	13.6	101.5	101.3	2.12
			$\sigma = 5.6$				Average 1.77
							$\frac{5.6}{1.77} = 3.16$

TABLE LVII

TEST OF PRESENCE OF FACTORS UNEQUALLY INFLUENCING SERVICE TENDENCIES OF CLASSROOMS

POPULATION	SD OF MEANS	AVERAGE OF SD'S OF MEANS	RATIO
X	3.3	1.67	1.98
Y	3.8	2.08	1.98
Z	5.6	1.77	3.16
XYZ	6.0	1.91	3.14

TABLE LVIII

PERCENTAGE DISTRIBUTIONS OF SERVICE REPUTATION SCORES OF SELFISH AND UNSELFISH BOYS AND GIRLS, BY POPULATION

Scores	Population X						Population Y					
	Boys		Girls		All		Boys		Girls		All	
	U	S	U	S	U	S	U	S	U	S	U	S
	*15**	*24*	*23*	*23*	*38*	*47*	*30*	*32*	*37*	*26*	*67*	*58*
High 36–44			4		3							
27–35							3		3		3	
18–26	7	4	4	4	5	4	10		3	4	6	2
9–17	33	17	22	29	26	23	30	19	30	19	30	19
0–8	47	48	70	17	61	32	37	25	59	65	49	43
− 9–1	7	17		21	3	19	10	31	5	8	7	21
− 18–10	7	9		17	3	13	7	9			3	5
− 27–19		4		12		9	3	3		4	1	3
Low − 36–28								12				7
Median					6.5						7.2	
Overlap						36						30
σ of U − S					3.1						3.7	

Scores	Population Z						Population XYZ					
	Boys		Girls		All		Boys		Girls		All	
	U	S	U	S	U	S	U	S	U	S	U	S
	25	*44*	*24*	*36*	*49*	*80*	*70*	*100*	*84*	*85*	*154*	*185*
36–44	4				2		1		1		1	
27–35							1		1		1	
18–26	4	2	4	8	4	5	7	2	4	6	5	4
9–17	20	11	50	17	35	14	27	18	33	18	31	18
0–8	52	32	33	39	43	35	44	26	55	49	50	37
− 9–1	12	27	12	14	12	21	10	27	6	13	8	21
− 18–10	8	14		3	4	9	7	13		4	3	9
− 27–19		9		19		14	1	8		11	1	9
− 36–28		5				2		6				3
Median					7.1		6.6		7.2		6.8	
Overlap						26		26		34		31
σ of U − S					4.5						6.7	

* The figures in italics in this and subsequent tables represent the number in each group.

TABLE LIX

PERCENTAGE DISTRIBUTIONS OF PHYSICAL CONDITION SCORES OF SELFISH AND UNSELFISH PUPILS OF POPULATION Y

SCORES	U	S
	35	*35*
High — 31–40	11	
— 21–30	11	9
— 11–20	9	6
— 1–10	23	14
9–0	11	29
19–10	14	23
29–20	6	17
39–30	6	3
Low 49–40	9	
Median	— 2.7	
Overlap		26
σ of U — S	1.2	

TABLE LX

PERCENTAGE DISTRIBUTIONS OF BURDICK CULTURE SCORES OF SELFISH AND UNSELFISH PUPILS, BY POPULATION

SCORES	POPULATION X		POPULATION Y		POPULATION Z		POPULATION XYZ	
	U	S	U	S	U	S	U	S
	39	*42*	*67*	*54*	*52*	*78*	*158*	*174*
High 240–275	46	29	9	6	4	3	17	10
204–239	23	14	16	9	17	5	18	9
168–203	23	26	19	17	15	13	19	17
132–167	8	17	21	26	33	22	22	22
96–131		12	28	31	25	35	20	28
60–95		2	6	11	6	20	4	13
Low 24–59						3		1
Median	234		159		153		176	
Overlap		32		38		30		32
σ of U — S	3.1		1.6		3.4		4.2	

TABLE LXI

PERCENTAGE DISTRIBUTIONS OF MOVIE ATTENDANCE OF SELFISH AND UNSELFISH PUPILS, BY POPULATION

AVERAGE WEEKLY ATTENDANCE	POPULATION X		POPULATION Y		POPULATION Z		POPULATION XYZ	
	U	S	U	S	U	S	U	S
	37	*48*	*61*	*54*	*48*	*65*	*144*	*165*
0	24	20	28	5	15	3	23	9
1	57	44	43	50	39	26	45	39
2	16	15	20	28	20	31	19	25
3	3	15	8	9	11	23	8	16
4		2		4	9	11	3	6
5 or more		4	2	4	6	6	3	5
Median	1.5		1.5		1.9		1.6	
Overlap		40		32		26		33
σ of U − S	2.1		2.7		1.7		4.0	

TABLE LXII

PERCENTAGE DISTRIBUTIONS OF INTELLIGENCE (CAVI SIGMA) SCORES OF SELFISH AND UNSELFISH PUPILS, BY POPULATION

SCORES	POPULATION X		POPULATION Y		POPULATION Z		POPULATION XYZ	
	U	S	U	S	U	S	U	S
	40	*47*	*73*	*57*	*48*	*80*	*161*	*184*
High 70–75		2						1
64–69	5	8	4	2			3	3
58–63	10	2	8	4	4	8	7	5
52–57	30	13	14	5	17	5	19	7
46–51	32	34	33	39	29	18	32	28
40–45	18	23	29	30	27	38	25	32
34–39	2	13	8	7	13	28	8	17
28–33	2	4	3	14	10	4	5	7
Low 22–27			1				1	
Median	51.2		48.7		46.0		48.1	
Overlap		30		39		31		34
σ of U − S	1.5		2.2		1.0		2.8	

TABLE LXIII

PERCENTAGE DISTRIBUTIONS OF EMOTIONAL STABILITY SCORES OF SELFISH AND UNSELFISH PUPILS, BY POPULATION

SCORES	POPULATION X		POPULATION Y		POPULATION Z		POPULATION XYZ	
	U	S	U	S	U	S	U	S
	38	*43*	*65*	*58*	*43*	*77*	*146*	*178*
High 4–0	16	7	9	14	5	4	10	8
9–5	24	19	26	12	12	10	21	13
14–10	21	16	17	14	2	9	14	12
19–15	13	26	12	29	23	22	16	25
24–20	10	19	17	9	19	14	16	13
29–25	8	2	8	7	19	18	11	11
34–30	5	7	5	3	16	12	8	8
39–35	3	5	3	9	5	3	3	5
44–40			3	3		1	1	2
49–45						4		2
Low 54–50						3		1
Median	11.4		13.4		21.4		15.6	
Overlap		34		38		52		41
σ of U − S	1.3		0.8		0.5		2.0	

TABLE LXIV

PERCENTAGE DISTRIBUTIONS OF AGES OF SELFISH AND UNSELFISH PUPILS, BY POPULATION

AGE	POPULATION X		POPULATION Y		POPULATION Z		POPULATION XYZ	
	U	S	U	S	U	S	U	S
	41	*50*	*74*	*60*	*55*	*86*	*170*	*196*
8		2	3			1	1	1
9	15	6	11	2	2		9	2
10	22	16	11	15	13	9	14	13
11	29	36	18	25	26	12	23	22
12	19	22	23	18	22	22	22	21
13	15	10	22	17	33	33	23	22
14		6	8	13	4	14	5	12
15			3	8	2	9	2	7
16		2	3				1	.5
17				2				.5
Median	11.5		12.3		12.4		12.1	
Overlap		44		49		34		41
σ of S − U	1.4		1.3		2.8		3.0	

2. REFERRED TO IN BOOK TWO

TABLE LXV

MEDIAN PERSISTENCE SCORE FOR EACH AGE LEVEL — POPULATIONS X, Y, AND Z

AGE	POPULATION X		POPULATION Y		POPULATION Z	
	N	Median	N	Median	N	Median
17			1			
16	1		4	180		
15	2	80	8	240	14	150
14	5	250	24	200	26	190
13	17	222	40	187	78	188
12	40	190	56	147	58	187
11	60	142	47	104	47	179
10	38	140	42	98	19	170
9	16	75	22	100	1	
Total		155.7		144.9		182.0

TABLE LXVI

MEDIAN INHIBITION SCORE FOR EACH AGE LEVEL — POPULATIONS X, Y, AND Z

AGE	POPULATION X		POPULATION Y		POPULATION Z	
	N	Median	N	Median	N	Median
17						
16	1		3	26.5		
15	1		9	25.8	9	21.2
14	6	15.5	19	22.9	23	20.6
13	16	25.0	40	23.2	66	22.5
12	40	24.3	54	22.8	62	24.0
11	60	23.1	44	21.2	50	22.2
10	36	23.0	39	22.1	20	18.0
9	12	22.3	25	24.1	1	
Total		23.2		22.7		22.2

TABLE LXVI (*Continued*)

POPULATION Y

AGE	BOYS		GIRLS	
	N	Median	N	Median
17	1			
16	2	26.0	1	
15	4	25.0	5	27.5
14	8	22.5	11	26.2
13	19	21.7	21	23.8
12	25	22.5	29	22.9
11	24	20.3	20	21.6
10	22	24.2	17	20.5
9	16	24.2	9	23.5
Total		22.3		22.8

TABLE LXVII

TOTAL PERSISTENCE SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	27	98.5	93.6	60.2	89	104.5	104.6	64.6	35	152.9	161.8	78.9
6	60	104.1	110.0	61.8	53	113.5	106.8	75.4	36	239.1	229.2	83.0
7	51	179.1	184.2	54.9	67	179.1	177.6	60.6	90	166.0	174.3	78.6
8	39	201.0	209.8	61.0	39	209.1	212.6	59.7	75	193.6	174.9	76.8
5 and 6	87	101.8	104.8	61.8	142	108.5	105.4	68.8	71	205.3	195.9	87.8
7 and 8	90	191.2	195.3	59.1	106	190.0	190.5	62.6	165	181.5	174.6	77.8
All	177	155.7	151.9	75.4	248	144.9	141.8	78.4	236	182.0	181.0	81.4

TABLE LXVIII

STORY RESISTANCE SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	34	2.8	2.92		97	2.3	2.45		41	2.6	2.75	
6	64	3.2	3.20		67	2.7	2.92		47	2.9	2.96	
7	78	3.6	3.62		72	3.2	3.40		131	2.7	2.89	
8	43	3.6	3.85		40	3.7	3.65		89	3.6	3.71	
5 and 6	98	3.1	3.10	1.19	164	2.4	2.64	1.48	88	2.8	2.80	1.36
7 and 8	121	3.6	3.70	1.02	112	3.4	3.48	1.02	220	3.2	3.36	1.18
All	219	3.4	3.43	1.14	276	2.9	3.01	1.36	308	3.1	3.12	1.24

TABLE LXIX

RING AND CROSS SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	36	3.8	11.65		99	15.3	18.30		45	24.6	29.15	
6	68	18.2	21.54		68	12.9	15.74		41	48.1	50.90	
7	80	27.9	30.25		73	24.4	30.90		120	32.5	37.28	
8	45	46.4	44.40		42	42.4	42.38		91	47.5	47.76	
5 and 6	104	15.0	18.10	17.25	167	14.1	17.26	13.88	86	35.0	39.52	18.79
7 and 8	125	34.7	35.35	19.30	115	28.4	35.09	18.08	211	38.3	41.80	20.45
All	229	23.2	27.52	20.38	282	18.9	23.65	18.10	297	38.8	40.95	20.10

TABLE LXX

MAGIC SQUARE SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	37	14.2	16.00		96	7.7	10.32		44	22.9	31.55	
6	65	21.4	23.42		67	18.3	19.22		46	60.0	50.10	
7	78	25.0	24.29		72	24.7	26.73		132	17.7	19.70	
8	43	29.8	32.40		44	24.4	26.69		88	22.6	25.33	
5 and 6	102	18.4	20.75	10.75	163	16.4	13.98	10.95	90	45.0	41.03	21.22
7 and 8	121	27.2	27.85	11.30	116	24.6	26.71	12.99	220	20.5	21.95	12.66
All	223	23.4	24.32	11.58	279	17.5	19.25	13.45	310	22.0	28.00	17.95

TABLE LXXI

PERSISTENCE FOR SELF SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	35	−2.1	0.4		94	5.4	7.0		46	−8.6	−9.1	
6	66	−9.1	−10.6		67	0.4	1.0		47	−9.3	−10.7	
7	76	10.5	10.4		75	5.4	4.6		128	9.2	14.8	
8	42	6.0	8.3		40	9.0	9.7		85	−15.0	−15.1	
5 and 6	101	−6.4	−6.8	28.7	161	3.7	4.5	21.8	93	−8.9	−9.9	30.6
7 and 8	118	9.0	9.7	20.0	115	7.0	6.4	20.0	213	2.4	2.8	39.7
All	219	3.6	2.1	25.8	276	4.9	5.3	21.9	306	−1.5	−1.0	37.6

TABLE LXXII

PERSISTENCE FOR CLASS SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	36	−10.0	−14.4		94	−2.1	−8.9		46	−1.4	−5.2	
6	66	−24.0	−27.0		67	−9.7	−16.8		46	−16.0	−18.0	
7	76	−3.1	−5.1		75	7.9	0.5		134	−3.3	−11.2	
8	42	−8.6	−9.0		40	0.0	−2.0		82	−13.3	−20.2	
5 and 6	102	−21.3	−22.5	26.5	161	−5.8	−12.2	29.3	92	−7.3	−11.6	37.6
7 and 8	118	−5.0	−6.5	29.6	115	2.2	−0.4	21.9	216	−7.9	−14.6	40.7
All	220	−10.6	−14.1	29.4	276	−2.7	−7.3	27.2	308	−7.7	−14.5	42.4

TABLE LXXIII

TOTAL INHIBITION SCORES BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	28	22.2	21.7	4.7	83	22.7	22.3	4.0	35	17.6	18.7	4.9
6	52	23.3	23.1	3.3	53	20.8	20.3	4.6	38	24.4	24.0	4.5
7	58	23.5	22.4	5.2	67	22.6	22.3	3.9	96	22.0	22.1	4.9
8	30	25.0	23.0	5.5	37	26.4	26.1	2.7	63	24.1	22.9	5.0
5 and 6	80	23.0	22.6	3.9	136	21.7	21.5	4.4	73	21.2	21.5	5.4
7 and 8	88	24.0	22.6	5.3	104	24.2	23.7	3.9	159	22.3	22.4	5.0
All	168	23.2	22.6	4.7	240	22.7	22.5	4.3	232	22.2	22.1	5.1

TABLE LXXIV

TOTAL INHIBITION SCORES WITHOUT THE PICTURE TEST, BY GRADE — POPULATIONS X, Y, AND Z

GRADE	POPULATION X				POPULATION Y				POPULATION Z			
	N	Median	Mean	SD	N	Median	Mean	SD	N	Median	Mean	SD
5	29	12.6	13.12	2.80	89	14.4	13.72	2.38	37	8.1	8.61	4.26
6	55	13.4	13.26	2.59	57	11.1	10.78	2.94	40	12.5	12.00	3.75
7	61	12.1	11.81	4.10	70	12.1	12.41	2.71	114	10.8	10.74	3.71
8	35	11.8	12.21	4.43	37	15.3	15.07	2.28	64	12.9	12.59	4.26
5 and 6	84	13.0	13.21	2.68	146	12.6	12.58	2.98	77	10.1	10.37	4.35
7 and 8	96	12.0	11.96	4.22	107	14.2	13.33	2.86	178	11.6	11.40	4.02
All	180	12.7	12.50	3.63	253	12.8	12.90	2.94	255	11.2	11.09	4.15

TABLE LXXV

Persistence Medians for Age and Grade

Age		Population X		Population Y		Population Z			
		N	Median	N	Median	N	Median		
Grade 5	9	12	173	22	100				
	10	19	115	29	107	16	170		
	11	5	55	20	93	12	115		
	12			9	110	5	150		
	13			3	130				
	14			3					
Grade 6	9	4	100						
	10	18	148	12	80	3	170		
	11	42	140	13	67	16	270		
	12	5	130	15	90	11	225		
	13			6	150	5	110		
	14			5	190				
	15			1	210	2	120		
Grade 7	11	8	160	14	158	17	150		
	12	11	185	24	175	26	170		
	13	3	170	16	181	36	184		
	14			9	205	8	230		
	15	2	80	4	175	5	165		
Grade 8	11	4	160			2	280		
	12	22	210	8	200	16	220		
	13	14	227	14	233	34	185		
	14	3	250	7	247	19	165		
	15			4	210	6	160		
	16			3	130				
	17			2	160				
								Total	
								N	Median
Accelerated . . .		8	130	56	124	38	189	102	148
Normal		146	156	138	140	180	174	464	158
Retarded		13	180	45	157	31	167	89	164

TABLE LXXVI

PERSISTENCE FOR SELF MEDIANS OF ACCELERATED, NORMAL, AND RETARDED PUPILS — POPULATION Z

GRADE	ACCELERATED		NORMAL		RETARDED	
	N	Median	N	Median	N	Median
5	22	− 11.4	13	− 5.0	10	− 6.7
6	3	− 45.0	33	− 5.0	7	− 25.0
7	22	0.5	102	+ 2.0	7	− 5.0
8	3	− 27.5	74	− 10.0	7	− 77.5

TABLE LXXVII

INHIBITION MEDIANS (THREE TESTS ONLY) BY AGE AND GRADE

AGE		POPULATION X		POPULATION Y		POPULATION Z	
		N	Median	N	Median	N	Median
Grade 5	9	11	11.3	24	14.5	1	8.0
	10	21	12.7	26	14.3	18	8.6
	11	4	13.5	19	12.8	13	3.5
	12			8	15.3	3	3.5
	13, 14			10	13.0	2	
Grade 6	9	3	15.3				
	10	13	14.3	13	9.3	3	13.5
	11	38	12.3	15	10.8	14	12.5
	12	6	12.5	17	11.5	13	10.8
	13, 14, 15			11	11.8	7	14.5
Grade 7	11	14	15.3	14	11.8	23	11.8
	12	19	12.5	27	12.3	33	11.5
	13	2	14.0	17	11.8	39	9.8
	14, 15, 16			15	13.5	18	10.0
Grade 8	12	17	12.5	8	14.2	17	14.4
	13	16	14.2	12	14.3	29	12.9
	14	6	7.0	7	14.7	16	10.5
	15, 16, 17	2	9.5	8	15.8	2	4.5
Accelerated		3	15.3	59	12.7	43	10.2
Normal		149	13.0	140	12.7	175	10.7
Retarded		20	10.9	52	14.0	32	9.3

TABLE LXXVIII

SEX DIFFERENCES ON THE STORY RESISTANCE TEST AND THE RING AND CROSS TEST

POPULATION	STORY RESISTANCE TEST				RING AND CROSS TEST			
	N	Mean	SD	G–B	N	Mean	SD	G–B
X Boys	114	3.36	1.18	+ .15	122	28.4	20.6	– 1.8
X Girls	105	3.51	1.08		107	26.6	19.9	
Y Boys	139	2.90	1.53	+ .22	144	23.8	19.3	– 0.4
Y Girls	137	3.12	1.18		138	23.4	16.9	
Z Boys	161	2.92	1.32	+ .42	160	41.1	21.1	– 0.3
Z Girls	147	3.34	1.17		137	40.8	18.8	
XYZ Boys	414	3.04	1.36	+ .27	426	31.6	21.7	– 1.1
XYZ Girls	389	3.31	1.16		382	30.5	20.0	

Girls – Boys = 3.0 σ Boys–Girls = .75 σ

TABLE LXXIX

SEX DIFFERENCES ON THE MAGIC SQUARE TEST

POPULATION	N	MEAN	SD	G–B
X Boys	120	24.0	12.2	+ .7
X Girls	103	24.7	10.8	
Y Boys	146	16.5	12.2	+ 5.8
Y Girls	133	22.3	14.0	
Z Boys	165	25.8	16.9	+ 4.8
Z Girls	145	30.6	18.7	
XYZ Boys	431	22.1	14.8	+ 3.0
XYZ Girls	381	25.1	15.7	

Girls–Boys = 2.8 σ

TABLE LXXX

SEX DIFFERENCES ON THE PERSISTENCE FOR SELF AND FOR CLASS TESTS

POPULATION		PERSISTENCE FOR SELF				PERSISTENCE FOR CLASS			
		N	Mean*	SD	G−B	N	Mean†	SD	G−B
X	Boys	111	+ 6.4	27.8	− 8.7	112	− 9.5	31.2	− 9.0
	Girls	108	− 2.3	22.7		108	− 18.5	26.8	
Y	Boys	138	+ 4.0	19.7	+ 2.6	138	− 11.7	29.6	+ 8.9
	Girls	138	+ 6.6	22.2		138	− 2.8	23.2	
Z	Boys	159	− 3.7	35.9	+ 5.6	160	− 13.9	39.3	− 1.2
	Girls	147	+ 1.9	39.1		148	− 15.1	45.4	
XYZ	Boys	408	+ 1.1	29.5	+ 0.7	410	− 12.0	34.2	+ 0.3
	Girls	393	+ 1.8	30.6		394	− 11.7	34.8	

Girls − Boys = .33 σ Girls − Boys = .14 σ

* The scores range from −100 to +150. † The scores range from −150 to +100.

TABLE LXXXI

SEX DIFFERENCES IN TOTAL PERSISTENCE SCORE AND IN REPUTATION FOR PERSISTENCE

POPULATION		TOTAL PERSISTENCE SCORE				REPUTATION FOR PERSISTENCE			
		N	Mean	SD	G−B	N	Mean*	SD	G−B
X	Boys	93	160.3	75.1	− 17.7	121	1.33	6.44	1.94
	Girls	84	142.6	74.7		110	3.27	6.04	
Y	Boys	121	131.7	79.4	+ 19.7	149	− 0.62	5.00	4.02
	Girls	127	151.4	76.5		142	3.40	4.76	
Z	Boys	121	169.4	78.3	+ 23.8	121	0.0	6.52	3.18
	Girls	115	193.2	83.2		112	3.18	6.30	
XYZ	Boys	335	153.0	79.5	+ 10.7	391	0.18	5.82	3.12
	Girls	326	163.7	81.5		364	3.30	5.58	

Population X, Boys − Girls = 1.6 σ
Population Y, Girls − Boys = 2.0 σ
Population Z, Girls − Boys = 2.3 σ
Population XYZ, Girls − Boys = 1.7 σ

Girls − Boys = 7.6 σ

* The scores range from −9 to +15.

TABLE LXXXII

Sex Differences on the Picture Inhibition Test

Population	N	Mean*	SD	G−B	$\frac{G-B}{\sigma D}$
X { Boys . .	121	− 5.97	14.56	+ 4.19	2.19
X { Girls . .	105	− 1.78	14.14		
Y { Boys . .	144	− 4.62	15.94	+ 1.26	0.68
Y { Girls . .	139	− 3.36	15.40		
Z { Boys . .	165	− 11.90	15.64	+ 16.54	8.89
Z { Girls . .	136	+ 4.64	16.38		
XYZ { Boys . .	430	− 2.95	15.72	+ 2.89	2.60
XYZ { Girls . .	380	− 0.06	15.84		

* The scores range about from −50 to +50.

TABLE LXXXIII

Sex Differences on the Puzzle Manipulation Test

Population	N	Mean	SD	G−B	% Inhibiting
X { Boys . .	112	5.45	2.04	+ .68	31
X { Girls . .	97	6.13	1.96		53
Y { Boys . .	142	6.17	1.63	+ .17	44
Y { Girls . .	130	6.34	1.39		40
Z { Boys . .	150	4.86	2.32	+ 1.00	30
Z { Girls . .	135	5.86	1.84		38
XYZ { Boys . .	404	5.48	2.32	+ .71	35.1
XYZ { Girls . .	362	6.19	1.45		42.6

Girls − Boys = 5.5 σ

TABLE LXXXIV

SEX DIFFERENCES ON THE SAFE MANIPULATION TEST

POPULATION		N	MEAN	SD	G–B	% INHIBITING
X	Boys	117	5.44	1.32	+ .40	40
	Girls	104	5.84	1.02		62
Y	Boys	145	5.69	1.20	+ .26	59
	Girls	137	5.95	1.09		71
Z	Boys	164	4.38	1.73	+ .73	30
	Girls	147	5.11	1.62		50
XYZ	Boys	426	5.12	1.57	+ .48	45
	Girls	388	5.60	1.35		64

Girls − Boys = 4.8 σ

TABLE LXXXV

SEX DIFFERENCES ON THE STORY INHIBITION TEST AND TOTAL INHIBITION SCORE

STORY INHIBITION TEST				TOTAL INHIBITION SCORE				
POPULATION		N	% Inhibiting	N	Mean	SD	G–B	$\frac{D}{\sigma}$
X	Boys	113	21	94	21.60	4.66	+ 2.3	3.3
	Girls	100	36	74	23.89	4.36		
Y	Boys	138	13	122	21.98	4.39	+ 1.0	1.7
	Girls	137	18	118	22.94	4.20		
Z	Boys	161	21	120	20.70	4.79	+ 3.1	4.5
	Girls	146	36	112	23.60	5.07		
XYZ	Boys	412	18.4	336	21.41	4.64	+ 2.0	5.5
	Girls	383	29.5	304	23.41	4.59		

Girls − Boys = 3.7 σ

TABLE LXXXVI

SEX DIFFERENCES IN REPUTATION FOR INHIBITION

POPULATION		N	MEAN*	SD	G–B	% REPUTED INHIBITING
X	Boys	117	2.40	9.83	+ 1.57	65
	Girls	108	3.97	8.72		77
Y	Boys	143	.12	11.04	+ 5.49	54
	Girls	133	5.61	7.95		73
Z	Boys	160	1.44	10.95	+ 2.64	58
	Girls	147	4.08	8.07		75
XYZ	Boys	420	1.26	10.71	+ 3.30	58
	Girls	388	4.56	8.31		75

Girls − Boys = 5.0 σ

*The scores range from − 28.5 to + 28.5.

TABLE LXXXVII

POPULATION DIFFERENCES IN TOTAL PERSISTENCE SCORE

POPULATION	N	MEDIAN	MEAN	SD
X	177	155.7	151.9	75.4
Y	248	144.9	141.7	78.4
Z	241	182.0	181.1	81.4

X − Y = 1.35 σ
Z − X = 3.78 σ
Z − Y = 5.45 σ

TABLE LXXXVIII

POPULATION DIFFERENCES ON THE STORY RESISTANCE TEST

POPULATION	N	MEDIAN	MEAN	SD
X	219	3.38	3.43	1.14
Y	276	2.92	3.01	1.36
Z	308	3.09	3.12	1.24

TABLE LXXXIX

Population Differences on the Ring and Cross Test

Population	N	Median	Mean	SD
X	229	23.2	27.52	20.28
Y	282	18.9	23.65	18.10
Z	297	38.3	40.95	20.10

TABLE XC

Population Differences on the Magic Square Test

Population	N	Median	Mean	SD
X	223	23.4	24.32	11.58
Y	279	17.5	19.25	13.45
Z	310	22.2	28.00	17.95

TABLE XCI

Population Differences on the Persistence for Self Test

Population	N	Median	Mean	SD
X	219	+ 3.6	+ 2.1	25.81
Y	276	+ 5.0	+ 5.3	21.90
Z	306	− 1.5	− 1.0	37.60

TABLE XCII

Population Differences on the Persistence for Class Test

Population	N	Median	Mean	SD
X	220	− 10.0	− 13.95	29.4
Y	276	− 2.7	− 7.30	27.2
Z	308	− 8.6	− 14.50	42.4

TABLE XCIII

Population Differences in Total Inhibition Score

Population	N	Median	Mean	SD
X	168	23.2	22.6	4.7
Y	240	22.7	22.4	4.3
Z	232	22.2	22.1	5.1

$X - Y = .44\,\sigma$ $X - Z = 1.01\,\sigma$ $Y - Z = .69\,\sigma$

TABLE XCIV

Population Differences in Puzzles + Safes + Stories

Population	N	Median	Mean	SD
X	180	12.7	12.5	3.6
Y	253	12.8	12.9	2.9
Z	255	11.2	11.1	4.1

TABLE XCV

Population Differences on the Picture Inhibition Test

Population	N	Median	Mean	SD
X	226	− 2.7	− 4.02	14.5
Y	283	− 3.4	− 4.00	15.7
Z	301	+ 2.7	+ 2.28	16.1

$Z - X = 4.9\,\sigma$

TABLE XCVI

Population Differences on the Puzzle Manipulation Test

Population	N	Median	Mean	SD	% Inhibiting
X	209	6.5	5.76	2.04	42
Y	272	6.7	6.25	1.53	42
Z	285	6.0	5.34	2.15	34

TABLE XCVII

Population Differences on the Safe Manipulation Test

Population	N	Median	Mean	SD	% Inhibiting
X	221	6.1	5.63	1.20	55
Y	282	6.2	5.82	1.16	65
Z	311	4.8	4.73	1.72	39

TABLE XCVIII

Population Differences on the Story Inhibition Test

Population	N	% Inhibiting
X	213	28.2
Y	275	15.6
Z	307	28.0

TABLE XCIX

National Differences in Total Persistence Score

Nationality of Parents	Population X		Population Y		Population Z		Population XYZ	
	N	Median	N	Median	N	Median	N	Median
Foreign . .	24	213	19	183	114	184	157	188
Mixed . . .	15	150	21	135	23	155	59	146
U. S. . . .	112	144	186	137	73	194	371	149
Italian . . .	7	205	5	150	40	200	52	195
Irish . . .	2	200	2	130	17	185	21	183

TABLE C

National Differences on Separate Persistence Tests

Puzzle Tests

Nationality of Parents	Population X		Population Y		Population Z		Population XYZ	
	N	Median	N	Median	N	Median	N	Median
Foreign . .	25	52.5	21	47.5	123	63.2	169	59.1
Mixed . . .	18	52.5	21	27.5	23	66.2	62	47.5
U. S. . . .	117	45.0	198	35.9	67	65.4	382	43.5
Italian . . .	5	32.5	6	35.0	44	65.0	55	58.1
Irish . . .	2	40.0	2	32.5	17	61.2	21	51.2

Persistence for Self and for Class Tests

Foreign . .	27	− 3.7	22	− 6.7	143	− 8.4	192	− 7.2
Mixed . . .	20	− 15.0	22	− 6.7	28	− 5.7	70	− 7.2
U. S. . . .	123	− 13.0	198	0.0	78	− 2.9	399	− 5.2
Italian . . .	7	− 10.0	5	− 10.0	58	− 6.7	70	− 7.5
Irish . . .	2	+ 20.0	2	+ 10.0	20	− 5.7	24	− 2.5

Story Resistance Test

Foreign . .	27	3.75	23	3.05	151	3.17	201	3.23
Mixed . . .	20	3.17	25	3.42	28	2.86	73	3.10
U. S. . . .	132	3.35	198	2.89	82	3.04	412	3.09
Italian . . .	7	4.25	5	3.38	62	3.19	74	3.27
Irish . . .	2	5.25	2	3.50	30	2.20	34	3.33

TABLE CI

National Differences in Total Inhibition Score

Nationality of Parents	Population X		Population Y		Population Z		Population XYZ	
	N	Median	N	Median	N	Median	N	Median
Foreign . .	19	25.5	19	24.5	137	21.9	175	22.6
Mixed . . .	18	25.0	21	23.8	22	20.5	61	23.5
U. S. . . .	110	22.8	178	22.2	62	23.7	350	22.6
Italian . . .	6	24.0	3	24.5	48	21.5	57	21.9
Irish	1		1		15	23.2	17	21.5

TABLE CII

National Differences on Separate Inhibition Tests

Picture Inhibition Test

Nationality of Parents	Population X		Population Y		Population Z		Population XYZ	
	N	Median	N	Median	N	Median	N	Median
Foreign . .	27	+ 2.5	22	+ 0.8	121	+ 5.6	170	+ 4.0
Mixed . . .	22	+ 3.6	23	− 2.5	20	− 3.3	65	− 0.4
U. S. . . .	135	− 3.7	207	− 2.8	74	0.0	416	− 2.5
Italian . . .	8	0.0	6	+ 2.5	52	+ 6.3	66	+ 4.4
Irish . . .	2	− 12.5	2	− 2.5	17	+ 6.9	21	+ 5.5

Puzzles + Safes + Stories — Population XYZ

Nationality of Parents	N	Median
Foreign	173	11.2
Mixed	61	12.5
U. S.	377	12.7
Italian	66	10.8
Irish	20	9.2

TABLE CIII

Classroom Persistence Scores

Population X

Class	Grade	N	Mean	SD	Boys		Girls	
					N	Median	N	Median
H	5	27	93.6	60.2	16	121.0	11	57.3
I	5 and 6	29	91.8	65.0	20	83.5	9	72.3
J	6	31	127.1	53.4	9	158.5	22	113.5
K	6 and 7	32	179.1	47.7	18	196.0	14	166.0
L	7	19	193.1	63.9	12	188.5	7	173.3
M	8	39	209.9	61.1	18	192.2	21	207.2

Population Y

N	5	29	93.3	70.7	16	53.5	13	102.2
O	5	30	131.6	56.9	14	156.0	16	136.0
P	5	30	88.5	56.7	19	88.5	11	83.5
Q	6	24	138.2	71.0	9	117.8	15	162.2
R	6	29	81.0	68.7	15	78.5	14	61.0
S	7	30	165.0	58.1	12	158.5	18	158.5
T	7	37	188.0	60.5	16	196.0	21	177.2
U	8	14	209.0	69.6	8	173.5	6	241.0
V	8	25	214.1	53.4	12	203.5	13	233.5

Population Z

A	5	35	161.9	78.9	17	117.2	18	203.5
B	6	36	229.2	83.0	19	231.6	17	263.5
C	7	37	164.9	74.9	17	124.8	20	196.0
D	7	22	159.6	92.7	13	143.5	9	143.5
E	7	31	196.2	66.0	17	182.9	14	196.0
F	8	35	198.2	57.6	17	203.5	18	216.0
G	8	40	154.5	85.4	21	173.5	19	154.8

TABLE CIV

Classroom Inhibition Scores

Population X

Class	Grade	N	Mean	SD	Boys		Girls	
					N	Median	N	Median
H	5	28	21.7	4.7	17	20.8	11	22.8
I	5 and 6	28	23.0	2.6	20	23.5	8	22.2
J	6	24	23.2	3.9	10	23.0	14	23.8
K	6 and 7	23	20.6	5.6	14	20.5	9	22.2
L	7	35	23.6	4.5	17	24.2	18	25.0
M	8	30	23.0	5.5	16	19.5	14	27.5

Population Y

N	5	24	21.6	4.6	14	24.0	10	22.0
O	5	29	23.1	3.9	12	24.5	17	22.2
P	5	30	22.1	3.4	22	23.0	8	21.5
Q	6	23	20.2	4.5	13	20.5	10	21.0
R	6	30	20.4	4.7	17	21.1	13	19.5
S	7	32	22.3	4.3	14	23.0	18	23.0
T	7	35	22.3	3.5	14	20.0	21	22.8
U	8	12	27.5	1.9	5	25.5	7	27.8
V	8	25	25.4	2.7	11	25.2	14	26.5

Population Z

A	5	35	18.7	4.9	16	16.5	19	18.2
B	6	38	24.0	4.5	20	21.5	18	27.5
C	7	36	23.6	4.4	17	22.2	19	25.5
D	7	38	20.0	4.5	23	20.4	15	19.8
E	7	22	23.0	5.1	11	20.5	11	25.8
F	8	36	24.4	4.0	18	25.5	18	24.4
G	8	27	20.9	5.6	15	20.5	12	21.5

TABLE CV

PER CENT OF OVERLAP OF THE NONPERSISTENT ON THE PERSISTENT GROUP — POPULATIONS X, Y, AND Z

VARIABLE	PER CENT OVERLAP			
	X	Y	Z	XYZ
Health		34.7		
Age	42.8	26.6	39.7	36.5
Reputation	26.2	44.8	27.7	40.0
Resistance to suggestion	42.0	36.1	48.9	40.4
Intelligence	41.8	45.7	42.8	41.6
Physical condition	39.2	38.6	50.8	44.4
Emotional stability	31.8	57.6	39.8	44.6
Cultural level	59.2	51.0	34.2	50.0

TABLE CVI

PER CENT OF OVERLAP OF THE NONINHIBITED ON THE INHIBITED GROUP — POPULATIONS X, Y, AND Z

VARIABLE	PER CENT OVERLAP			
	X	Y	Z	XYZ
Reputation	39.0	33.3	25.8	32.8
Health		41.2		
Emotional stability	50.7	44.1	30.4	41.5
Physical condition	51.5	12.0	56.5	43.9
Age	48.1	37.1	43.2	44.7
Cultural level	40.5	51.3	34.4	45.5
Intelligence	33.9	66.1	42.8	46.8
Resistance to suggestion	42.8	38.5	64.3	53.9

TABLE CVII

PROBABILITY OF OCCURRENCE OF DIFFERENCES BETWEEN NONPERSISTENT AND PERSISTENT GROUPS — POPULATIONS X, Y, AND Z

VARIABLE	CHANCES IN 1000			
	X	Y	Z	XYZ
Reputation	945	726	971	992
Resistance to suggestion	945	964	864	992
Nail biting		986		
Intelligence	841	540	977	971
Emotional stability	986	579	618	945
Health		933		
Physical condition	618	726	758	618
Age	691	955	758	726
Cultural level	758	655	964	500

TABLE CVIII

PROBABILITY OF OCCURRENCE OF DIFFERENCES BETWEEN NONINHIBITED AND INHIBITED GROUPS — POPULATIONS X, Y, AND Z

VARIABLE	CHANCES IN 1000			
	X	Y	Z	XYZ
Reputation	816	1000 *	1000 ‡	1000 §
Intelligence	964	955	903	977
Emotional stability		726	997	977
Physical condition	618	1000 †	997	933
Nail biting		816		
Age	994	977	864	691
Cultural level	841	618	841	758
Health		691		
Resistance to suggestion	618	816	758	540

*3.6 σ †3.8 σ ‡4.6 σ §5.4 σ

APPENDIX D

COMPUTATION OF CAVI SIGMA SCORES

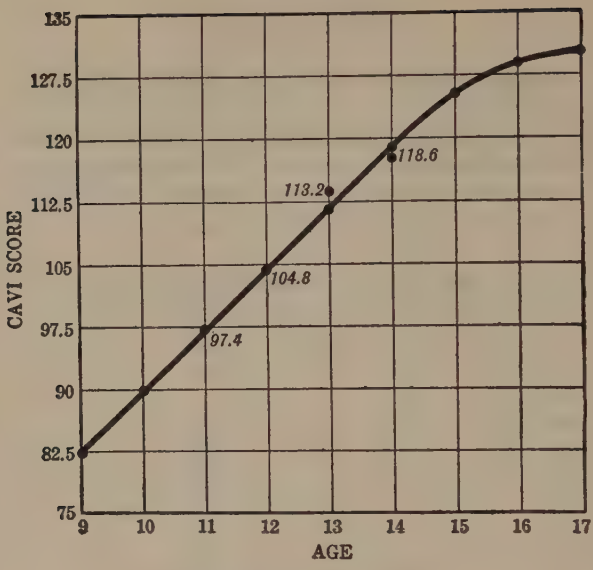
In order to have a practicable measure of intelligence which was relatively free from the influence of age, we first found the means and SD's of the CAVI achievement scores for each complete chronological year for all the cases we had. These distributions are shown in the accompanying table.

DISTRIBUTIONS OF CAVI ACHIEVEMENT SCORES BY AGE

SCORES	11	12	13	14	11–14
190–199				1	1
180–189	1		5	5	11
170–179		9	10	12	31
160–169	6	7	12	11	36
150–159	6	22	16	18	62
140–149	12	30	23	17	82
130–139	34	44	26	30	134
120–129	60	71	42	31	204
110–119	70	60	48	42	220
100–109	88	93	46	34	261
90–99	118	106	47	37	308
80–89	94	79	21	22	216
70–79	81	57	32	13	183
60–69	49	35	10	8	102
50–59	23	13	8		44
40–49	5	7	1	1	14
30–39	1		1		2
20–29	1			1	2
N	649	633	348	283	1913
M	97.40	104.80	113.20	118.60	105.90
SD	24.10	26.76	29.85	29.66	27.89

With these materials in hand, a curve was drawn for which the means for ages 11 and 12 determined the slope and from which the means for younger and older groups could be computed. This curve is shown in the figure on the next page.

In calculating the means and SD's of ages beyond the known limit, account was taken of the fact that there is a difference of opinion regarding the shape of the curve of mental ability. We did not wish to be dogmatic



AGE CURVE OF CAVI SCORES

in insisting on the straightness of the curve beyond the ages for which we had a reasonably accurate measure and therefore arbitrarily dropped the curve for ages beyond 14. The following table shows the conventional adaptations made in the means and SD's in this projection of the curve.

CONVENTIONAL ADAPTATION OF MEANS AND SD'S OF CAVI SCORES FOR AGES 7 TO 17

AGE	MEAN	SD
7	67.5	20
8	75.0	20
9	82.5	21
10	90.0	22
11	97.5	24
12	105.0	27
13	112.5	30
14	120.0	31
15	125.0	32
16	129.0	33
17	130.0	33

On the basis of these figures we computed a sigma score for each child, dividing his deviation from the mean of his age by the SD for his age.

A table was made for reading off these sigma scores from the original scores and at the same time translating them into convenient whole numbers, with an average or zero sigma value at 50. A section of the key for age 11 follows:

TRANSMUTATION KEY FOR CAVI SIGMA SCORES

SCORE	SIGMA DEVIATION FROM MEAN	KEY VALUE
90	− .31	47
91	− .27	48
92	− .23	48
93	− .19	49
94	− .14	49
95	− .10	49
96	− .06	50
97	− .02	50
98	+ .02	50
99	+ .06	50
100	+ .10	51
101	+ .14	51
102	+ .19	51
103	+ .23	52

APPENDIX E

DIRECTIONS FOR ADMINISTERING SELF-CONTROL TESTS AS GIVEN IN POPULATIONS X, Y, AND Z

These tests were given as a part of an extensive test program which will be described more fully in Volume III. For purposes of discussion in the present volume they have been lifted bodily from this larger program, and the directions for administration are printed here in the order in which the self-control tests were administered but without reference to the other tests which accompanied or paralleled them.

1. Picture Inhibition Test. *Form 1–b.*

" Now we must turn to some practice work in simple arithmetic. We must practice starting and stopping together; so when I say,

' Go,' fill out the blanks at the top of the page for name, school, grade, and age. Never mind the others. Ready? Go."

Allow 2 minutes.

" These exercises are arithmetic speed work and do not count on your monthly grades. The scores will not be given to the teacher. They will have nothing to do with your standing or promotion. This work is practice work in adding. It will merely help you to do your arithmetic better.

" Now open the folder to test 1. All pencils up. Spread it open so that pages 2 and 3 are both up. Look at the directions. Read them. If you finish page 2 before time is called, follow the arrow to page 3 and keep right on adding. Do you understand? Pencils up. Ready? Go."

Allow 2 minutes and then say: " Stop. Pencils up. Turn to test 2, page 4. Do these just as you did the others. Ready? Go."

Collect papers while assistant passes next test, face up, saying, " Do not touch."

Form 1–a.

" When I say, ' Go,' write your name, school, grade, and date only on the lines provided. Ready? Go.

" Pencils up. Now before you turn the page let me explain that on the following pages are some pictures, stories, and other interesting things, as well as some more addition problems. You will have time to look at them for only a short time. If you want to make a good score, follow directions. Start adding promptly when I say, ' Go,' and stop when I say, ' Stop.' Turn to page 4. Here is a story about a boy who got lost in a dark forest. (Examiner reads first three lines aloud.) Keep pencils up. You may have a few seconds to read."

Allow 20 seconds and say: " All ready now. If you want to make a good score, do the additions of test 1 as fast as you can. Begin at the left. Go."

Allow 2 minutes and say: " Stop. Pencils up. Turn at once to page 6. This is a garage. There are twelve mistakes in the picture, and on the right is a puzzle picture. There is a third boy hidden in the picture. Keep your pencils up while looking at the picture and rest your hand."

Allow 45 seconds and say: " All ready now. If you want to make a good score, do the additions of test 2 as fast as you can. Begin at the left as before. Go."

Allow 2 minutes and say: "Stop. Pencils up. Turn at once to page 8. Keep your pencils up. A few seconds for the headlines while resting."

Allow 10 seconds and say: "All ready for the additions of test 3. Begin at the left. Go."

Allow 2 minutes and say: "Stop. Pencils up. Turn at once to page 10. Here are some riddles to amuse yourself with for a moment. The answers are on page 11. Keep your pencils up."

Allow 20 seconds and say: "All ready for the additions of test 4. Begin at the left. Go."

Allow 2 minutes and say: "Stop. Pencils up. Turn at once to page 12. Here are some nuts to crack; the answers are on page 13. Keep your pencils up."

Allow 20 seconds and say: "All ready now for the additions of test 5. Begin at the left. Go."

Allow 2 minutes and say: "Stop. Pencils up. Turn at once to page 14. I guess you can see what this is. Keep your pencils up."

Allow 8 seconds and say: "All ready now for the additions at the left. Go."

Allow 2 minutes and say: "Stop now. Pencils up. Close the booklets. Pass them quickly to the front."

2. Persistence for Self and for Class Tests.

Directions identical with Efficiency Coöperation Test, Appendix B.

3. Story Resistance Test.

"Keep these papers face down." (Pass the practice exercise S–R 1 for the Story Resistance face down.)

"Turn the sheet over and write your name, date, school, and grade at the top. Now read the directions. Do you understand? Put a short up and down line after each word. Look at the sample. (Explain in detail.) All right. Go ahead."

Allow 3 minutes and say, "Stop."

Collect the papers.

"How many of you like stories? I want to find out how well you like them. I am going to read you a story. Listen carefully." (Read the Lucille story to the point indicated.)

"How many would like to know how the story ends? I'm going to let each of you find out for yourself. The sheets I shall pass out have the ending printed on them, but the words are run together just

the way they were in the exercise you did a moment ago. You can best read the story by drawing a short line after each word just as you did before. Leave the papers alone till all are passed."

Pass the ending to the Lucille story (a single sheet) — S–R 2.

" Now turn over the page and put your name and date on it. Read the directions; if you care about the story badly enough to find the words, go ahead and draw the little upright lines after each word." (Note the time here or start your stop watch.)

Allow 2 minutes and say: " As soon as you feel you would rather quit than finish the story you may do so and I will give you another story to read. If you would rather quit, you may bring your paper up and I will give you another story to read — the first part of which is easy reading."

When anyone brings up a paper, put down the time on his paper and give him " Dark Terror " — S–R 3. Give "Unseen Traps" last — S–R 4.

At 5-minute intervals repeat the invitations to stop.

4. Magic Square (and Word) Test.

Pass out P–M 1 (Magic Number Squares), saying, " Keep this sheet face down until I ask you to turn it up."

When all are passed say, " Turn over the sheet and fill in all the blanks at the top." (Allow 2 minutes.)

Say: " Now read the directions silently as I read them aloud. (Read.) You understand, it makes no difference *what* numbers you use, but you must not use any one number more than one time. If you use 2, you may use 2 but once. The numbers you do use may add up to any number you please — up and down, across, and across corners — but it must always be the *same* amount in the three columns down, three rows across, and across the corners." (If necessary, make a blackboard diagram.)

Say: " This is not a speed test. Any questions? Ready? Go." (Note clock time at start.)

At the end of 2 minutes say, " Any time you would rather quit trying to make a magic number square, bring your sheet up and I will give you another sort of puzzle to work on."

Repeat invitation to stop at 5-minute intervals.

When a child brings up P–M 1, give him P–M 2, the Magic Word test.

Always note clock time on P–M 1 and P–M 2 when they are returned.

When a child brings up P–M 2, give him in succession:

The Blacksmith's Chain
The Squares and Circles
The Sum of Fifteen Puzzle
Letter Counting

As you give the last group of puzzles, always say, " Be sure to read the directions carefully." Do not note time on these last puzzles.

Allow 75 minutes as maximum time. If all are finished with P–M 1 and P–M 2 and the group grows restless after 50 or 60 minutes, call in all the papers.

5. Cross and Ring Puzzle.

" To-day we will do a puzzle. You may have your choice of two puzzles to try — this Japanese Wooden Cross or these Chinese Rings. To solve the Wooden Cross puzzle, you must take these pieces and put them together to make a cross like this (hold up one all put together). To solve the Chinese Rings, you must take the rings off the bar without straining or bending them and then put them on again the way they were before. Now as we pass them each one say which he wants to work with." (Pass puzzles and note time.)

Every 5 minutes say, " If any of you are tired of this and want to quit, you may come and get the other puzzle." Later say also, " If any of you want to quit the second puzzle, I have some other kinds to work on."

When a puzzle is brought up, note the time opposite the pupil's name. Allow 75 to 80 minutes, or less if all have quit on the second puzzle.

6. Puzzle Manipulation.

This was administered along with a deception test. The essential directions were as follows:

Materials.

1. Fifteen puzzle in standard arrangement (for the deception test)
2. Puzzle peg box with sundry puzzles, etc., accurately arranged according to master set
3. Three puzzle fillers, one marked with a numeral corresponding to the number on a peg box

4. Letter counting sheets
5. Score sheets

Note to Testers. Items 1, 2, and one of 3, marked with number of box, are all passed out at the same time and fastened together so that subject gets a peg box and filler having the same number. Item 5 is passed out afterwards.

First Day. "I am going to pass out all our things together. The large box (hold up one) has several things in it which we would rather you would not disturb. If you get these things disarranged, it simply means a lot of work for someone to put them right again. We will use these boxes presently. Just let things rest on your desk until I give the directions."

Pass out items 1, 2, and the first filler, and then item 5.

Then follow directions for the deception test. After this is finished, the examiner says: "Now, next, we will take the Arrows puzzle (filler No. 1). Write your name, etc. at the top and read the directions. Do you understand? Go ahead."

Pass the second filler.

At the end of 3 minutes say: "Stop. Be sure the sheet is filled in at the top. Lay this sheet to one side and I will collect it. Go to work on the Wall puzzle (second filler)."

Collect the first filler, passing the third at the same time.

At the end of 3 minutes repeat the directions and collect the second filler, passing the letter counting sheets at the same time.

At the end of 3 minutes repeat the directions and say: "Go to work on the letter counting exercises. Be sure your name is at the top."

At the end of 3 minutes say: "Stop. Tear off the sheet or sheets on which you have been working and I will collect them. Go ahead on the next sheet."

Ditto at the end of two more 3-minute intervals.

Then say: "Now take your score sheet once more and place it at the right. Put the box in front of you so that the side where it says T is away from you, at the top. Don't touch anything in the box. Notice that most of the holes have pegs in them. Now take your pencil and mark on your paper a letter X on the little circles to show just where the pegs are, and leave the other circles blank. (Explain on blackboard.) Don't move the things to see where the holes are. Just put an X where the pegs are. Go ahead."

Allow 3 minutes and collect the papers, saying, "Please do not disturb

the things in the box." Then collect the boxes *with great care not to drop them.*

7. Safe Manipulation.

This was administered along with a deception test. The essential directions were as follows:

Materials.

1. Puzzle Peg game, in boxes (for deception test)
2. Safes, all combinations turned to $6\frac{1}{2}$
3. Word making sheets, numbered W–M 1, 2, and 3
4. Letter counting sheets
5. Score sheets

Note to Testers. Items 1, 2, 3, and 4 are all passed out at the same time. Item 5 is passed out afterward.

First say: "To-day we are to have several things to work with, and I am going to have them all passed out at once. We shall not have time to work with all the puzzles which you had on your desk the other day, but we shall have a chance to solve the Puzzle Peg, as well as to see if we can find the secret of the combination to this safe. Please do not touch any of the things which are put on your desk until all are ready to begin. If you touch anything now, it will give you an unfair advantage over the others. So please let everything alone until it is time to begin."

Pass out items 1, 2, 3 (1 sheet), then item 5.

Say: "Now turn first to the Puzzle Peg game in the box and let the safe strictly alone. The object of this game is" — etc., continuing with this deception test. When this is completed, the examiner says: "Before trying to open the safes we will turn to some other things for a moment. Look at the first sheet, headed 'Word Making.' Do you understand what you are to do? All right. Go ahead."

Pass out the second W–M.

At the end of 3 minutes say: "Stop. Be sure you have this sheet filled in at the top. Lay this sheet to one side and I will collect it. Go to work on the next 'Word making' sheet." (Collect W–M 1 and pass W–M 3 at the same time.) *Mark safe dial reading on every sheet as collected.*

At the end of 3 minutes repeat the directions and collect W–M 2, passing the letter counting sheets at the same time.

At the end of 3 minutes say: "Stop. Be sure you have this sheet filled in at the top. Lay this sheet to one side and I will collect it.

Go to work on the letter counting exercises. Be sure your name is on the top. If you finish sheet 1 before time is called, go to sheet 2."

At the end of 3 minutes say: "Stop. Tear off the sheet or sheets on which you have been working and I will collect them. You may go ahead on the next sheet."

Ditto at the end of two more 3-minute intervals.

Then say: "Now turn to the safe. If you can find the combination, you may have the money which is inside."

At the end of 4 minutes say: "Now turn to your score sheet. If you found the combination, write it in the proper place. If you got the safe open, put an X after the words 'I did get the safe open.' If you did not get the safe open, place an X after the words 'I did not get the safe open.' Be sure your name is on the score sheet."

Allow time; collect score sheets and everything left on desks.

8. Story Inhibition Test.

Say: "I am going to pass each of you a story. Do nothing with it until all are passed." Pass "Shipwreck," *face down.*

"Turn the paper over and fill in the spaces at the top. Now we'll read the story and see what it's all about." (Examiner reads to bottom of page 2.)

"Now I want to see how clever you are at inventing a good ending for this story. On the blank sheet you may write an ending, telling how you think the story turns out. It would be a fine thing for you to think it all out for yourself; but if you feel that you *have* to see what really happens, you may break open the sheet and read the rest of the story." (Repeat the last sentence.)

Allow 3 minutes, passing out "Mystery Box," face down, saying, "Do not touch this until I give directions, please."

At the end of the second minute say, "You have just one more minute." Then say, "Stop now and fill in the blanks at the top of the first sheet of this second story."

"This time we will do the story as it was first planned. It is to see how accurately you can do certain kinds of work. The score will not count on your grades. First we will read the words. I'll read aloud."

Read to bottom of page 1 and say, "Now notice the directions at the bottom of the page." Read them and say, "Decide which you will do — finish the story or see how well you can do on the test."

Allow 3 minutes and collect papers.

In addition to these tests the Candy test was given in population C. The procedure was the same as for the Safe Manipulation test. The open boxes of candy were passed out at the beginning of a long session of testing, and as the examiner collected papers he indicated on each paper just what candy had disappeared from each pupil's box. There were four checks of this sort. The test began as follows:

"As a reward for taking so many tests we are going to treat everyone to a small box of candy. This candy is yours, but we ask that you leave it in the boxes until we have finished the exercises for to-day. If everyone is chewing candy, it adds to the confusion. But we will pass it out now anyway."

Pass the boxes, lids off, repeating: "Please leave the candy in the boxes till after the exercises to-day. It adds to the confusion to have everyone eating candy now."

APPENDIX F

RESULTS OF SELF-CONTROL TESTING IN POPULATION C

Six tests were tried out in population C:

Story Inhibition
Puzzle Manipulation
Safe Manipulation
Picture Inhibition
Ruggles Distraction
Candy Inhibition

As these techniques offer interesting possibilities, we are recording certain facts about them which may be of some assistance to others.

RELIABILITY

The Story Inhibition Tests. These two tests employed the same technique — that of reading a story to the climax and giving the pupil the option of tearing open the seal and reading the ending or doing something else. The two stories were different, and the alternatives to finishing them were different. In one story it was to guess how

the story ended, while in the other it was to go back and count letters in order to make a good score. The two tests were given on two different days, each in a battery with other tests. The sequence is given in Chapter III.

Only two scores were used, 0 or 1, according to whether the pupil inhibited or did not. In the following table, which gives results for both stories, " Yes " means the pupil opened the paper; " No " means he inhibited. Thus 107 pupils opened the sheet on both stories and 122 did not open it on either. The figure at the bottom is the per cent (73) which showed the same behavior on both tests.

TABLE A

RESULTS OF STORY INHIBITION TEST — POPULATION C, TEN GROUPS

		STORY I		
		Yes	No	N
STORY II	Yes	107	32	139
	No	53	122	175
	N	160	154	314

73%

This indicates considerable consistency in conduct in view of the fact that the two tests were quite different. We have estimated the coefficient of reliability of this test to be about .50.*

The Manipulation Tests. In the first manipulation test the object was a small toy safe, and the reason given for not touching it was that it would be unfair in view of the fact that a contest was to follow. In the other test, the Puzzle Manipulation, the objects were small puzzles set in among a group of pegs, and the reason given for not

* No one of the various formulas for fourfold tables quite applies. We have, however, computed various measures of relationship. The r computed from the formula $r = \text{cosine } \pi \text{ U}$ (where U is the proportion of pairs of unlike signs to total number of pairs) is .66; but ϕ is .46, Q .45, and C .45. The tetrachoric r is not strictly applicable, since the table is not split at the median both ways. However, it is nearly so vertically, and the horizontal split is about 45% and 55%, which Kelley says should give good results. See Kelley, *Statistical Method*, p. 257. But since the result is so much higher than the other formulas give and also since the other formulas agree so closely, we are led to doubt whether the true reliability coefficient is around .45 or around .65. If these two figures may be taken as limits, we shall not be far wrong if we say that the reliability coefficient is .50.

touching was that it would make a lot of work for the examiner to get the boxes back in order. Here again the technique is the same, but the objects are different and the inhibiting resistance is different.

In Table B we present the data according to the number who yielded or did not yield to the temptation to manipulate. Thus out of 302 cases, 59 did not touch either the safe or the puzzles, 117 touched or manipulated both. The per cent exhibiting similar behavior is 58 for all cases combined.

TABLE B

RESULTS OF SAFES AND PUZZLES TESTS — POPULATION C

		TEN GROUPS PUZZLES		
		Yes	No	N
SAFES	Yes	117	18	135
	No	108	59	167
	N	225	77	302
		58%		

		SIX GROUPS PUZZLES		
		Yes	No	N
SAFES	Yes	101	17	118
	No	18	55	73
	N	119	72	191
		82%		

In groups 6A1, 6B2, 7A3, and 8B1 a large percentage manipulated the puzzles and a small percentage the safes. This is partially explained by the way the Safes test was administered in these groups. These were the first four groups tested, and the examiner checked on the safes by writing or making a mark on a paper in the presence of the pupils. It became evident to them that they were being checked up, and this probably tended to cause them to "shy off" from touching the safes. We later changed the method of checking, making it much less obvious that any note was being made of whether or not they handled the safes.

If we omit from consideration these four groups and deal with the remaining six who were tested with the later and improved technique of checking, the results are more consistent. Now we find that almost exactly the same number manipulated the safes as manipulated the puzzles, 82% of them showing similar behavior. The coefficient of correlation between the puzzles and safes, computed from these limited data, is .61.

Since both of these tests admit of more detailed scoring than the bare dichotomous "did" or "didn't" manipulate, we were able to secure data that give more than two categories for each test. As

a first approximation we scored the Puzzles test by counting one for each puzzle misplaced in the box (for detailed description see Chapter III). These raw scores ran from zero to 7. The safes were checked three times and we counted one for each time the examiner noted a change on the dial. The scores run from zero to 3. Omitting groups 6A1, 6B2, 7A3, and 8B1, the correlation between these more detailed scores is .50.

The Visual Distraction Tests. These two tests, the Picture Inhibition and the Ruggles, are based on the double testing technique. (See *Studies in Deceit.*) The score on each is the difference between the number of examples added when distracting visual stimuli are on the pages before the pupil and the score when the page contains only the examples to be added. The two tests as used in population C are the same in so far as each requires the work of adding. But in the Picture test the type of example is of this sort: $\begin{matrix}2 & 4 & 5\\ \underline{3} & \underline{6} & \underline{8}\end{matrix}$, while in the Ruggles test the examples are the more difficult type: 5 8 6 3 4 7 2 9 = ? Another important difference is that the distractions of the Picture test are always placed at the top of the page with the examples below, while in the Ruggles test the distractions are all mixed in with the examples. The reader should turn to Chapter III and compare the appearance of these two tests to comprehend fully how different they are.

The correlation between the raw difference scores of these two tests is + .35. This coefficient probably does not represent the true relationship between the influence of visual distraction on these tests. It is too low because of the presence of errors of type I, described in *Studies in Deceit*, Book Two, Chapter VI.* Correcting for this type of error, we get a corrected r of .42.

While this r is high enough for a single pair of tests, we feel that it does not represent the true reliability of this sort of material. Since the Picture test has six separate sections, or exercises, we have correlated the odd-numbered exercises against the even-numbered and found

* Very briefly, the difficulty is that the *difference* scores represent partly the influence of the distraction and partly a host of other unknown factors, which are ordinarily lumped together and called errors of measurement of adding ability. While we do not have at hand all the data that would enable us to correct the correlation for the influence of such errors, as we did in *Studies in Deceit*, yet we can approximate it.

that the test as a whole has a predicted reliability of about .90. This simply means that, if a pupil stops work on one set of examples to look at one kind of picture, puzzle, or whatever it may be, he will probably do so again on another set, even though the distracting stimuli are not identical.

The Candy Inhibition Test. This had no other test to match it; so no measure of reliability was secured. However, we found that the correlation between the amount of candy missing from the box after the first two checks correlated over .90 with the amount missing at the end of the fourth check. But this cannot be regarded as a measure of the reliability of the test. Unfortunately we could not repeat it with some other kind of sweets, such as Eskimo pies.

Reliability of the Entire Battery. The reliability of the entire procedure was determined in the following way. Two sets of tests and parts of tests were made up. The first consisted of:

1. The odd-numbered exercises of the Picture Inhibition test
2. The first Story test
3. The Safe test
4. The Ruggles test

The second consisted of:

1. The even-numbered Picture Inhibition exercises
2. The second Story
3. The Puzzle test
4. The Candy test

The correlation of set I with set II is .628. From this the Spearman prophecy formula shows that the entire test has a reliability of .77.

VALIDITY

How well these tests individually or collectively measure inhibition or self-control may be determined theoretically from the intercorrelations of the test situations. This method of determining validity is described in detail in *Studies in Deceit*, Book One, Chapter V, and we shall not repeat the discussion here.

We have intercorrelated these tests in two ways. First we have combined the two tests in each of the first three types of situations into a single test score for that type of situation. Each of these are intercorrelated and each is correlated with the Candy test. The results are given in Table C.

TABLE C

INTER-*r*'s OF INHIBITION TESTS — POPULATION C

TEST	OBJECTS	STORIES	CANDY
Pictures *	.34	.32	.24
Objects		.20	.15
Stories			.21
Candy			

* These *r*'s are all based on the Picture test alone. The Ruggles test was dropped out.

The average of these inter-*r*'s is .243, which is slightly higher than the average inter-*r*'s of all the deception tests (.227, see *Studies in Deceit*, Book Two, p. 125). Basing our computations on this figure, we find that these four situations have a theoretical reliability of .56 and a theoretical validity of $\sqrt{.56}$, or .748. In order to secure a reliability of .90 and a validity of .95, we should need to measure inhibition in 28 similar situations all told.

There is still another way of dealing with these data, and that is to regard each class score as a unit and use class medians or class percentages as measures. This of course reduces the number of cases to 10. However, there is considerable justification for this because it appears that each class behaved as a unit. An examination of the number of each group who failed to inhibit reveals the fact that something like an " all or none " law is at work. In the Stories, Safes, Puzzles, and Candy tests, where it was possible for any pupil to observe the behavior of his classmates, we find that the per cents who fail to inhibit are either very large or very small. If these per cents for these tests are plotted, we get a distinctly bimodal distribution.

The tendency is for either around 20% to 30% to fail or 90% to 100%. The middle of the scale, 40% to 60%, has no cases at all. In view of this it would appear that the safest statistical procedure is to regard each group as a unit and make all computations from group percentages or medians. This would make possible also the correction of the individual *r*'s for heterogeneity.

Table D shows the tests ranked by median amount of inhibition shown by each group on six tests. (Since the correlation between

the ranks for the Story tests was .90, we combined the rankings.) The inter-*r*'s are shown in Table E. Two of them are negative and two are zero. They average .275, which yields a predicted reliability of .695 and a validity of .835.

TABLE D

SUMMARY OF GROUP RANKS BY MEDIAN SCORES ON INHIBITION TESTS — POPULATION C

GROUP	STORIES	SAFES	PUZZLES	CANDY	PICTURES	RUGGLES
4B4	1	2	1	3.5	2	4
5A2	8	5	2.5	2	3	5
5B2	9	7	2.5	6	6	1.5
6A1	10	6	4	10	8	1.5
6B2	5.5	3	8	7	4	8
7A1	5.5	10	10	5	10	10
7A3	2	1	5	3.5	5	6
7B1	4	8	6	8	1	7
8A1	7	9	7	9	7	9
8B1	3	4	9	1	9	3

TABLE E

INTERCORRELATIONS OF RANKS OF GROUPS ON INHIBITION TESTS — POPULATION C

TEST	SAFES	PUZZLES	CANDY	PICTURES	RUGGLES
Stories	.53	− .11	.52	.36	− .22
Safes		.36	.49	.37	.36
Puzzles			.10	.57	.61
Candy				.02	.16
Pictures					.01

Average, .275
Predicted reliability, .695
Predicted validity, .835

In Table F we have combined the tests, putting the two Story tests together, the Manipulation tests together, and the two Picture tests

together. In each of these combinations we have averaged the ranking of the component tests. These inter-r's are somewhat higher than those in Table E, but they are less reliable. The average is .386, which gives a predicted reliability of .715 for the battery and a predicted validity of .845.

TABLE F

SUMMARY OF GROUP RANKS BY MEDIAN SCORES ON COMBINED INHIBITION TESTS — POPULATION C

CLASS	STORIES	MANIPULATION	CANDY	PICTURES
4B4	1	1	4	1
5A2	9	6	2	2
5B2	8	3	5	5
6A1	10	5	9	7
6B2	6	4	6	8
7A1	4	9	7	10
7A3	2	2	3	5
7B1	5	9	8	3
8A1	7	9	10	9
8B1	3	7	1	5

INTERCORRELATIONS

TEST	PICTURES	OBJECTS	STORIES
Candy	.56	.44	.40
Pictures		.45	.22
Objects			.25

Average, .386
Predicted reliability, .715
Predicted validity, .845

The application of Spearman's tetrad difference criterion to Table C yields tetrads of zero, indicating that these four situations hang together by a common factor. The correlations between the test situations and the general factor, whatever it is, are .68, .44, .48, and .35 for the Picture, Manipulation, Story, and Candy tests, respectively.

APPENDIX G

1. STANDARDIZATION OF THE PICTURE INHIBITION TEST MATERIAL

The Picture Inhibition test was standardized on a population in the same city with populations X, Y, and Z. The tests without distracting pictures were given in the same way as later when pictures were used. We used for our standardization testing one fifth-grade class, one sixth, two seventh, and one eighth, a group which corresponded closely in grade constitution to our populations X, Y, and Z.

It is required to know the correlation between the practice period, *i.e.*, the first four minutes of adding, and the test periods, *i.e.*, the last twelve minutes. The work was divided into units of two minutes each, and the required r is thus between the first two periods and the last six. In this population of 170 pupils this correlation is .963. If this should seem high, the reader is reminded of the fact that we are dealing here with a very narrow mental function and that correlations of .90 between two periods of two minutes each are not uncommon with this sort of material, especially when the range of talent runs from the fifth to the eighth grade.

In finding the mean gain of the test periods over the practice we divided the test score by 3, reducing all to a two-period, or four-minute, basis. For this population the mean number of problems added in the first two periods (*i.e.*, the first 4 minutes) is 163.17, and one-third of the mean of the last six periods (or 12 minutes) is 176.81, leaving a mean gain of 13.64 points, or problems added. The distribution of test gains and losses is given in Table A.

The SD of this distribution is 12.92 points and the mean 13.64. Thus 13.64 becomes the point of reference and 12.92 the unit by which the distracting influence of the pictures is measured when the pictures are used. The formula for computing the conventional Xi score for this test is:

$$\text{Inhibition Xi} = \frac{(P - \frac{1}{3}T) - 13.64}{12.92},$$

where P = practice score and T = total test score.

In order for the Xi to become − 3.00, there must be a loss on the tests of 26 points. If an individual's total Picture test score divided

by 3 is 26 points below his practice score, the chances are something like 999 in 1000 that he was distracted by the pictures. As a matter of fact, Table A shows no losses greater than 15.

TABLE A

DISTRIBUTION OF GAINS AND LOSSES OF THE STANDARD GROUP ON THE PICTURE INHIBITION MATERIAL

GROSS DIFFERENCES	STANDARD GROUP	
	N	%
+ 35 + 44	8	4.7
+ 25 + 34	26	15.3
+ 15 + 24	45	26.5
+ 5 + 14	45	26.5
− 5 + 4	33	19.4
− 15 − 6	13	7.6
	170	100.0

Mean, 13.64
σ, 12.92

2. STABILITY OF THE DERIVED PICTURE INHIBITION SCORES

It might occur to the reader that the superiority of Z on the Picture Inhibition test is apparent only and could be accounted for by the way in which the score is derived. The score is the difference between initial ability as measured by four minutes of adding and one-third of the work done in the twelve succeeding minutes, while distractions were present. If the Z children, because of lack of interest or language ability or adaptability to the new situation, did relatively more poorly on the practice test than the X children, it might be expected that as they proceeded and caught on to the task they would improve more rapidly than X and so show a larger difference score. As a matter of fact, however, they scored an average of 174 examples during the first four minutes of work as against 150 examples for X, so that the handicap, if there was any, was to the disadvantage of X, not of Z.

Furthermore, when the Persistence for Self and for Class tests were given later, after the pupils had had these twelve more minutes of

practice on the inhibition test, the first four minutes of this second test showed that the Z pupils had gained 7.9% in ability over their first four minutes on the earlier test and that the X pupils had gained 9.6% over their first four minutes. Consequently we must believe that Z was as near its zenith of ability as was X when the preliminary practice on the Picture test was done, and that the difference in rate of work under distraction represents a genuine difference between X and Z in distractibility.

A further check on this is afforded by the following comparison: The Persistence for Self and for Class tests involved twenty-four minutes of work. No distractions were presented. We may therefore compare the difference between the first four minutes on this test and the next twelve minutes with the corresponding difference between the four minutes of practice on the Picture Inhibition test and the twelve minutes of work done under distraction. Table B gives the relevant figures.

TABLE B

COMPARISON OF POPULATIONS X AND Z WITH RESPECT TO ABILITY TO GAIN ON A SPEED TEST WITH AND WITHOUT DISTRACTION

(The figures represent the average number of examples done in four minutes.)

	POPULATION X		POPULATION Z		RATIO
	Examples	Gain	Examples	Gain	
Practice period on Picture test	150.0		181.0		.830
		− 4.0		+ 2.3	
One-third distraction period on Picture test . .	146.0		183.3		.797
First four minutes on Persistence test	165.6		195.4		.848
		+ 2.1		− 2.5	
One-third next twelve minutes on Persistence test	167.7		192.9		.870

From this table it can be seen that X *lost* 4 examples under distraction and *gained* 2.1 without it, whereas Z *gained* 2.3 under distraction

and *lost* 2.5 without it. These differences may be compared by means of the four ratios of X to Z which are involved. The fall in the ratio of X to Z of .830 to .797 shows to what extent X is more distracted than Z, and the rise in ratio from .848 to .870 shows to what extent X overtakes Z when there is no distraction.

It need not be inferred that Z was actually stimulated by the pictures and other distractions to do better than if they had not been there, for our standard population, using material identical with the Picture test except for the absence of the pictures, showed a mean gain of 13.6 examples as against Z's gain of 2.3 under distraction.

APPENDIX H

DIRECTIONS FOR INHIBITION PARTIES AS GIVEN IN POPULATION C

Preparations. 1. Prepare identification badges of cloth or heavy paper with half the badges numbered from 1 to — in red and half from 1 to — in yellow.

2. Prepare simple score cards with spaces for the numbers 1 to — and several columns for each game, in which the judges will enter checks indicating failures to inhibit.

3. Secure and train enough judges so that in each game no one will need to watch more than four or five children. Have one judge to help select prize winners and prepare list of winners and prizes.

4. Have prizes and souvenirs ready for distribution.

5. Prepare class roll for listing identification numbers.

6. Prepare material for Noise Response game.

7. Secure an entertainer for the half of the group not engaged in test games.

8. Mark off chalk lines for Hopping Race and Crows and Cranes before the children report for the party.

9. Find out the name of the class teacher and write it in the proper place in the Funny Story game.

10. As pupils enter the room have their badges pinned on and the number of the badge and the word "Red" or "Yellow" placed opposite each child's name on the roll.

Plan of Party.

Examiner directs as follows: "Your principal has agreed that it will be a nice thing for us to show our appreciation to those pupils who have been taking our tests by giving them a party here in the gymnasium. During this next hour you are our guests at a party. There are to be games that will give each of you the chance to win a real prize. There is one thing, however, that we must ask of all of you before we go on with this plan. As we have only one hour, it will be impossible to carry out the plans for the party unless each of you is willing to do just as you are told and follow all instructions carefully. Will all who understand this please raise one hand? Remember, do just as you are are told, and the party will be a pleasant time for everybody.

"First, let's walk past that table and get the party badges; be sure and put your badge on so that the judges who give the prizes can see the number that wins. If your badge comes off, put it back on as soon as possible. Help each other to pin the badges on and then fall in double line here. Move out quickly. . . .

"Now, children, we are ready to enjoy the games. As there are too many of you to play the same game, we will divide you into two groups, the Reds and the Yellows. You (designating the dividing couple) will please move over near that wall, and Miss Blank will see that you are entertained. Please do as she tells you to avoid confusion. The rest of you remain with me."

Test Games.

1. Hopping Race.

"Now this group is to take part in a hopping race for a prize. Listen to the rules. Each of you now take a place on the chalk line here; one person to a space, as marked off, all standing straight and facing squarely in this direction (indicating). I am going to count slowly up to three, in this way (counting). When I say, 'Three,' you may immediately take one step with either foot you want to and then begin hopping, keeping in your own chalk lines. The one to reach that line (indicating) the most times out of five races wins the prize. Running or walking will not count. Do not start until the word 'Three' is actually spoken by me.

"Do you all understand? Are all ready? On the line! One, two, three! (You started too soon; you ran; this young lady wins the first heat.) On your lines quickly for the second race. On the line!

One, . . . two, three! Now, back on the starting line for the third race. On the line! One, two, . . . three!

"Now back in place for the fourth race. Give me your attention. This time no one is to start until I count to six. When I say the word 'Six,' it's time to start. All understand? On the line! One, two, three, four, five, six! Now get ready for the last race. On the line! One, two, three, four, five, six!

"The judges will award the prize to the winner at the end of the party."

2. *"Crows and Cranes."*

"The next game is very interesting. It is called 'Crows and Cranes.' Line up here, please, children, in double rank. That's right. Now you in the front rank are the crows and are to take places on this line. The other rank are the cranes and will be here. Look and be sure you remember who is just opposite you in the other line. Remain facing each other. When I say, 'Crows!' the crows start quickly to chase the cranes, each crow trying to touch his crane before he reaches the back line. If I say, 'Cranes!' the cranes chase the crows to their back line. Who are the crows? That's right; and you others are the cranes. Remember that crows and cranes do not talk in this game. Back on your lines for the next trial. All ready! On the lines! Cr-Cr-Cr-Crows! Now! Cr-Cr-Cr-Crows! Get ready again. All set! Cr-Cr-Cr-Cranes" — etc., six times.

3. *Funny Story Game.*

Both groups are consolidated for this game.

"Now, children, will you all form a circle by catching hold of hands? Drop hands now and sit down. The object of this next game is to see who can keep a straight face, no matter what I say to you. We will try it three different times. The judges will pick the one who best deserves the prize. Remember, don't move or even smile with your eyes. When I say, 'Hold,' the game is on. When I finish the joke or remark, remember this: the game is not over until I say, 'Let go.' Does everybody understand? Ready. Hold."

1. A son at school wrote to his father, "No mon, no fun, your son."
The father wrote back, "How sad, too bad, your dad."
After 2 seconds say, "Let go!"

2. Where is Times Square?" asked Miss Blank (substitute name of teacher).

" I'm standing on it," replied the traffic officer.

" Then no wonder I couldn't seem to find it," said Miss Blank.

After 2 seconds say, " Let go ! "

" All set for the third trial. Hold ! "

Wait, silent, 2 seconds, then say :

3. " Children, I don't mind your sitting there sleeping with your eyes open, but I certainly do object to that loud snoring."

After 2 seconds say, " Let go."

4. *Simon Says, " Arms Up."*

" Will you all now get up and form in a double line right here? We're going to play an old game, but it isn't so easy to do. In this game two prizes will be given, one for each line. The back line kindly turn around and walk four steps. Now every other one in both lines step forward one step. That's fine. This game is called, 'Simon says, " Arms up."' Listen carefully to the directions. When I say, 'Arms up,' don't move your arms the least bit. Leave them absolutely still at your sides. But when I say, 'Simon says, " Arms up,"' then raise your arms quickly over your head, this way (demonstrating). Don't lower your arms unless I say, 'Simon says, "Arms down."' If I say, 'Arms down,' don't move them ; just leave them up. Remember that Simon must tell you when to move your arms ; if he doesn't, then do not move them. Does everybody understand? All ready ! Simon says, 'Arms up' ; Simon says, 'Arms down' ; Arms up." Wait a second, for the judges to record errors, then say : " All who have arms up please put them down ; remember to wait until Simon says put them up. Simon says, 'Arms up' ; Arms down." Correct as above and continue, " Arms down." Again correct and continue : " Simon says, 'Arms down' ; Simon says, 'Arms up.' Simon says, 'Arms down.' "

5. *Noise Response Game.*

" Now, children, every party should have at least one thrill, shouldn't it? Well, the next game certainly has that ! Will you all catch hands and form a circle? That's right. Now drop hands and I'll tell you the rules of the game. The object of the game is for all of you to face outward — away from me — and then to see who can keep from turning this way or even looking this way when all sorts of noises are made here in the center of the circle. That is, you are not to turn, no matter what I do or say, except for one thing, and that is

this police whistle (showing it). Now, if you will listen to it, I'll blow it three times so that you can recognize it. (Demonstrate it.) The judges are going to try to pick the quickest in turning when this whistle blows; but of course the prize winners will also have to be the ones who do not turn on the wrong sounds. Two prizes are to be given in this game. Are all ready for the fun? You may practice turning quickly if you want to. That's good. Are you clear on the rules? If so, then please face outward and I'll bring in the noise machines (procures suit-case, containing clap-boards, sacks [paper], tin pan, siren, bottles [in basket, for breaking with hammer]). Now, all ready. Hold!!! The game is on."

Here alternate noises, as follows: (1) whistle, (2) whistle, (3) sack burst, (4) siren, (5) whistle, (6) whistle, (7) whistle, (8) help (sharp yell), (9) whistle, (10) clap-boards, (11) siren, (12) broken bottles, (13) whistle, (14) whistle. Pause between noises (1) and (2); (4), (5), (6), and (7) should follow each other in rapid succession; also (9) should quickly follow the yell (8).

6. *The Hall Test.*

Announce, " The game is over. Will you face this way?

" That's not always easy, is it? The next one surely is and you'll probably like it best, for now we are ready to march out into the hall, where the souvenirs of the party will be given to each of you. We hope you have enjoyed the games, and we want to give you a little present, not much, but just something to remember us by.

" Will you all fall in double rank again? That's right. Now all face that door. There's just one thing I must tell you. There are other classes at work in the classrooms near the hall out there, and we of course do not want to disturb them. So we shall ask you the second you pass out of the gymnasium door to stop all noise and talking and to remain quiet until you enter the door of the gymnasium again. Will you all do this for us? Thanks. Remember, quiet and no talking at all when you reach the door to pass into the hall to get your souvenirs; you may wait and talk about the whole thing after you return to this room. Please follow me by the table and back here, keeping your present places in line. Let's go."

Upon the return, prizes are awarded and the class then sent to its proper room.

NOTE. The Hall test was given only twice, as everyone was quiet.

INDEX OF TOPICS AND PROPER NAMES

Achievement Capacity test, 290
Administration of tests, self-control tests, 318–320; service tests, 19–35, 55–73
Age differences, in self-control, 361–369, 466; in service, 38–41, 120–125, 254, 265
Allport, F. H., 52
American Child Health Association, 160, 245, 385–386, 433
Amount of, helpfulness, 21–22, 25–28; self-sacrifice, 30–36; service, 271; unselfishness, 261
Associates, *see* Friends, Sociability
Attendance, *see* Motion pictures, School attendance, Sunday-school attendance

Background, *see* Community differences, Cultural background differences, Socio-economic differences
Bakwin, Ruth M., 245
Behavior scales, 42–46
Behavior tests, *see* Inhibition, Persistence, Service tests
Burdick Apperception test, 172, 175, 185–189, 267. *See also* Correlations, Cultural background differences
Burtt, H. E., 287

Cabot, Ella Lyman, 248
Candy Inhibition test, administration of, 316; reliability of, 541; scoring of, 317
Catholics, *see* Religion
Chapman, J. C., 290
Character as self-control, 283–284
Character education, 273–275, 452–454
Character Education Inquiry, Foreword, 3, 4
Character tests, *see* Inhibition, Persistence, Service tests
Charity, *see* Helpfulness, Self-sacrifice tests, Service
Check List, administration of, 91–93; description of, 92, 323–324; reliability of, 93–94, 326–327; scoring of, 93, 321–322. *See also* Reputation
Churches, *see* Religion
Classroom, *see* Groups, School morale
Cleanliness, and self-control, 433–434, 449; and service, 235–236, 269
Club membership, and self-control, 421–425, 449; and service, 224–228. *See also* System X, System Y
Coe, G. A., 49
Community differences, in helpfulness, 41; in miscellaneous factors, 169–177; in self-control, 364, 391–395, 448; in service, 169–180, 267
Competition, 47–49

Conclusions, 7–10, 263–273, 443–452
Conduct, specificity of, 246–248, 445–446
Conduct Record, description of, 94, 322; scoring of, 94–95, 322, 324–325
Conflict, 42–49
Coöperation, 47–49; of parents, 205–206; teaching of, 248–251; tests of, 49–52. *See also* Efficiency Coöperation test, Free Choice test, Helpfulness, Self-sacrifice tests, Service
Correlations between: age and school attendance, 209; age and self-sacrifice, 40; home background and self-sacrifice, 40; intelligence and self-sacrifice, 40; physical condition and self-sacrifice, 40; siblings and self-sacrifice, 40; suggestibility and self-sacrifice, 40; years in school and self-sacrifice, 40

Inhibition and: age, 365, 446; Burdick, 396, 448; deportment, 379, 447; emotional condition, 388, 389, 447; friends, 436–437; health score, 386; intelligence, 374, 379, 446; movie attendance, 427; persistence, 365; physical condition, 385, 447; reputation for inhibition, 336; school attendance, 411, 412; school marks, 375, 376, 379, 446; sex, 416; siblings, 409–410, 448; Sims, 396, 448; speed ability, 438–439; suggestibility, 429, 449; Sunday-school attendance, 420, 449

Persistence and: age, 363, 378, 446; Burdick, 396, 448; emotional condition, 387, 447; friends, 436–437; health score, 386; inhibition total, 365; intelligence, 372–373, 378, 446; physical condition, 385, 447; reputation for persistence, 331; school attendance, 411; school marks, 375, 379; sex, 416; siblings, 408–409, 448; Sims, 396, 448; speed ability, 438–439; suggestibility, 428–429, 449; Sunday-school attendance, 419, 449

Service and: age, 121, 123, 145, 265; Burdick, 185, 188, 267; Check Lists, 103–104; classmates, 215–216; cleanliness, 235; Conduct Record, 103–104; deportment, 143, 266; emotional condition, 165–167, 266–267; friends, 237–238; "Guess Who," 103–104; intelligence, 139, 145, 266; majorities, 239; physical condition, 160–161, 266; reputation, 103–108, 117; school attendance, 209; school marks, 143–145, 266; siblings, 202–205, 267; Sims, 187–188; sociability, 236; suggestibility, 235
Crane, A. L., 285
Criteria of testing procedures, 17–18. *See also* Validity
Cross and Ring test, 295–296, 533
Cultural background differences, in self-control, 396–397, 448; in service, 172–175, 185–189, 193–197, 267. *See also* Community differences

Deceit, 280–282
Denomination, *see* Religion
Deportment, 143, 176, 266, 379, 447
Dewey, John, 47
Directions for giving, self-control tests, 293–320, 529–537; service tests, 19–35, 55–77, 459–463

Distraction tests, 286–287. *See* Picture Inhibition test, Ruggles Distraction test
Distributions, 21, 22; Burdick scores, 173; "Guess Who" scores, 90; movie attendance, 231; Sims, 174; total individual persistence scores, 349; total inhibition reputation scores, 337; total inhibition scores, 338; total party scores, 360; total persistence scores, 391; total service reputation scores, 105; total service scores, 109, 178
Dodge, R., 285
Downey, J., 288
Drives, 71–72, 280–283

Economic factors, *see* Community differences, Cultural background differences, Occupational differences, Socio-economic differences
Efficiency Coöperation test, administration of, 68–70; illustration of, 69; directions for giving, 459–463; reliability of, 100, 103–104; scoring of, 70
Emotional condition, and self-control, 387–389, 447; and service, 162–168, 266–267
Envelopes test, administration of, 64–66; description of, 63–66; reliability of, 100; scoring of, 66–67; validity of, 103
Environment, and self-control, 390–406; and service, 169–199, 267. *See also* Community differences, Cultural background differences, Occupational Differences, Socio-economic differences

Family, *see* Parents, Sibling resemblance
Fernald, G. G., 290
Follett, M. P., 49
Free Choice test, administration of, 70–71; reliability of, 100; scoring of, 70
Friends, resemblance in self-control, 436–437, 450; resemblance in service, 237–239, 267. *See also* Sociability

Galton, F., 407
Games, party, *see* Party games
Gilliland, S. F., 288–289
Grade, *see* School grade
Group, differences in self-control, 451–452; differences in service, 212–214, 272; morale in self-control, 414–416, 417, 449; morale in service, 212–217, 268; resemblance in self-control, 416; resemblance in service, 214–216
Groups, contrasted, in self-control, 441–442; in service, 252–261
"Guess Who" test, 87–91, 325–326; administration of, 87–89; description of, 87–88; reliability of, 91; scoring of, 90–91; validity of, 91

Health, 160–161, 266, 385–386, 447
Helpfulness, difficulties of testing, 16–17; reliability of tests, 28–29; tests of, 14–41. *See also* Cooperation, Self-sacrifice tests, Service, Service tests
Heredity, *see* Sibling resemblance
Home influence, and self-control, 407–410, 448; and service, 200–207, 267–268

Ice Cream test, 34–35, 456
Individual differences, in self-control, 450–451; in service, 271–272

Individual tests of self-control, description of, 342–345; scoring of, 346–347; validity of, 349–352; value of, 347–352
Inhibition, as interest, 280–282; as self-control, 279; classroom tests of, 302–317, 529–531; individual tests of, 343–352; party tests of, 548–552; previous efforts to measure, 285–291; reputation for, 323–327; results of tests, 537–540. *See also* Self-control
Intelligence, of parents, 206–207, 268; scoring method, 527–529; and self-control, 371–374; and service, 40, 139–143
Intercorrelations of, helpfulness scores, 28; individual inhibition scores, 351; individual persistence scores, 348; inhibition classroom scores, 339, 340, 542, 543; party inhibition scores, 358; persistence classroom scores, 332, 442; reputation for inhibition scores, 327; reputation for service scores, 96; self-sacrifice scores, 37; service scores, 111
Interest as self-control, 280–283, 376–378
Irwin, Will, 42

Kelley, T. L., 116, 167, 246
Kerr, C. A., 248
Kropotkin, P., 42

Learning Exercises, 59–61

Majorities, influence of, on self-control, 439–440, 450; on service, 239–270
Maller, J. B., 67, 72, 250, 251
Manipulation tests, *see* Puzzle Manipulation tests
Marston, L. F., 289
Mathews, Ellen, 163
McCaskill, J., 51, 52, 55
McDougall, W., 280
Measurement, *see* Tests
Money Sacrifice test, 29–34, 456
Money Vote test, administration of, 56; reliability of, 100; scoring of, 57; validity of, 103
Moore, H. T., 286, 288
Moral Education, *see* Character education
Morale, *see* Group, School morale
Morgan, J. J. B., 289
Motion pictures, and self-control, 425–427, 449; and service, 228–232, 269
Motives, 5, 46–47, 50, 79–80, 241, 270
Müller, J., 50

Nail biting, 167, 388
Nationality, and self-control, 286, 397–401; and service, 189–194, 255, 267

Occupational differences, in self-control, 395–396; in service, 172, 175, 181–185, 267
Otis, Margaret, 233

Parents, coöperation of, 205–206, 268; intelligence of, 206–207, 268, 410, 417
Party games, inhibition tests, 49–50; administration of, 353–355, 548–552; description of, 355–356; reliability of, 356–357; results of, 358–360; validity of, 358
Persistence, as interest, 280–282; classroom tests of, 292–301, 317–320; in coöperation, 243–245; individual tests of, 342–348;

reputation for, 321–323, 330–333. *See also* Self-control
Persistence for Self and for Class tests, 300–301, administration of, 531; reliability of, 329–330. *See also* Efficiency Coöperation test
Physical ability test, 158–161
Physical condition, and self-control, 385–386, 447; and service, 158–162, 266. *See also* Health
Physical types, 245–246
Picture Inhibition test, administration of, 313–315, 529–531; illustration of, 315; reliability of, 334–335; scoring of, 316, 546–548; standardization of, 545–546
Picture Pasting test, 19–24
Populations tested, 6–7. *See also* Community differences
Portrait Matching device, 77–78, 86–87
Pupil Data Sheet, 463–466
Puzzle Manipulation tests, administration of, 305–306, 533–535; illustration of, 306; reliability of, 334; scoring of, 306–307
Puzzle Mastery tests, administration of, 295–297; description of, 295–297; facsimiles of, 298–299; scoring of, 296–300

Race, *see* Nationality
Ratings, *see* Check List, Conduct Record, "Guess Who" test, Portrait Matching device, Social Service Record
Reliability, of helpfulness tests, 28–29; of individual inhibition tests, 351; of inhibition tests, 333–335, 341, 537–544; of party inhibition tests, 356–357; of persistence tests, 328–330, 341; of reputation scores, 87, 91, 94, 96; of self-sacrifice tests, 37; of service tests, 100–101; of social service record, 77
Religion, in reputation, 404; in self-control, 401–405; in service, 194–199
Reputation for inhibition, 326
Reputation for persistence, 323
Reputation for self-control, 321–327; religious differences in, 404; sex differences in, 380–385
Reputation for service, 74–77; community differences in, 176; distribution of ratings, 105; relation to service tests, 107–117, 253–254; techniques for measuring, 75–95
Results of tests, self-control tests, 443–445, 456, 537–544; service tests, 21–24, 25–28, 30–34, 35–36, 456–459
Roback, A. A., 284
Ruggles Distraction test, administration of, 310–311; facsimile of, 308–311; scoring of, 311–313; standardization of, 312–313
Russell, Bertrand, 42

Safe Manipulation test, 305
Scaling of tests, 37–39, 455–459
Scholastic status (retardation-acceleration), and self-control, 368–371, 446; and service, 130–139, 266
School attendance, 40, 209–210, 411–412, 417, 449
School differences in self-control, 412–414, 417, 449
School grade, and self-control, 367–369, 446; and service, 125–130, 265–266
School influence, on self-control, 410–417, 448–449; on service, 208–217

School Kits test, administration of, 61–62; description of, 61–62; reliability of, 100; scoring of, 63; validity of, 100
School marks, and self-control, 375–376; and service, 143–144, 176, 266
School morale, and self-control, 414–416; and service, 208–217, 268
Self-control, analysis of, 279–284; and character, 283–284; conclusions concerning, 7–10
Self-control tests, *see* Inhibition, Persistence
Self-sacrifice tests, description of, 29–35; scaling of, 38–39, 455–459; validation of, 36. *See also* Coöperation, Helpfulness, Service
Self or Class test, 55, 103
Selfish and unselfish groups, comparison of, 252–259
Sengupta, N. N., 52
Service, analysis of, 14–16; conclusions concerning, 8–9; an object of study, 42–51
Service tests, administration of, 55–71; criteria of, 17–18; description of, 53–73; distribution of, 109; reliability of, 100; scoring of, 72, 101; validity of, 101–118
Sex differences in, classroom groups, 214–215; helpfulness, 22–23; reputation for self-control, 381–385; self-control tests, 380–385; self-sacrifice, 40; service tests, 146–157, 266
Sibling resemblance, in self-control, 408–410, 417, 448; in self-sacrifice, 40; in service, 201–205, 267
Sigma scores, 527–529
Signature, effects on scores, 89
Sims, V. M., 51, 172, 175, 181, 395
Sinha, C. P. N., 52
Sociability, in self-control, 435, 449–450; in service, 236, 269
Social Resistance test, 289
Social Service Record, 75–77
Socio-economic differences, in self-control, 396–397; in service, 172–175, 187–189, 193, 197, 267
Spearman, C., 246
Specificity, of self-control, 442, 445–446; of service, 246–248, 270
Speed-ability test, 437–439, 450
Standard procedures for testing, 17–18
Standardization of tests, 545–546
Story Inhibition tests, administration of, 302–303; description of, 302–303; illustration of, 304; reliability of, 334; scoring of, 303
Story Resistance tests, administration of, 292–295; illustration of, 294; reliability of, 328–329; scoring of, 294–295
Suggestibility, and self-control, 428–437, 449; and service, 233–235, 269
Summer camps, 228, 425
Sunday-school attendance, and self-control, 418–421, 449; and service, 219–224, 268
Symonds, P. M., 47
System X and service, 23–24, 224–257
System Y, and self-control, 421–424; and service, 225–227, 257

Tables, list of supplementary, 466–470
Teacher influence, *see* School influence
Teacher rating, *see* Check List, Conduct Record, Portrait Matching device, Reputation
Teaching morals, 10, 248–251, 270

Teams, 240–270
Techniques for testing, *see* Tests
Test situations, 53–54
Tests, *see* Helpfulness, Inhibition, Persistence, Self-sacrifice tests, Service tests
Theophrastus, 77
Thomas, W. I., 42
Thorndike, E. L., 171
Toy Making test, 24–26

Uhrbrock, R. S., 288
Unselfishness, 13–14, 259–261

Validity, 74–75; theoretical, 110–111; true, 112–117, 264. *See also tests by name*

Washburne, C., 51
Weinland, J., 243
Welty, R. E., 342
Whipple, G. M., 291
Woodworth-Mathews Personal Data Sheet, *see* Emotional condition
Woodworth, R. S., 163, 280
Word Building test, 290–291

Yule, G. U., 243

14977

The Evangelical
School of Theology

LC
268
.C75
1928a
v.2

v.2. 41,830

DATE DUE

UNITED THEOLOGICAL SEMINARY

UNITED THEOLOGICAL SEMINARY LIBRARY
3 0645 1000 9992 4